

GenCore version 5.1.6
Copyright (C) 1993 - 2004 CompuGen Ltd.

OM nucleic - nucleic search, using sw model
Run on: June 27, 2004, 03:11:33 ; Search time 4007.03 Seconds
(without alignments)
10708.573 Million cell updates/sec

Title: US-10-658-691-3
Perfect score: 990
Sequence: 1 atgtaccactcaaaattgta.....atttgacgcattttcttaa 990

Scoring table: IDENTITY NUC
Gapop 10.0 , Gapext 1.0

Searched: 3470272 seqs, 21671516995 residues
Total number of hits satisfying chosen parameters: 6940544

Minimum DB seq length: 0
Maximum DB seq length: 2000000000
Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : GenEmbl.*
1: gb.ba.*
2: gb.htg.*
3: gb.in.*
4: gb.om.*
5: gb.ov.*
6: gb.pat.*
7: gb.ph.*
8: gb.pl.*
9: gb.pr.*
10: gb.ro.*
11: gb.sts.*
12: gb.sy.*
13: gb.un.*
14: gb.vi.*
15: em.ba.*
16: em.fun.*
17: em.hum.*
18: em.in.*
19: em.mu.*
20: em.om.*
21: em.or.*
22: em.ov.*
23: em.pat.*
24: em.ph.*
25: em.pl.*
26: em.ro.*
27: em.sts.*
28: em.un.*
29: em.vi.*
30: em.htg.hum.*
31: em.htg.in.*
32: em.htg.other.*
33: em.htg.mus.*
34: em.htg.pln.*
35: em.htg.rod.*
36: em.htg.man.*
37: em.htg.vrt.*
38: em.sy.*
39: em.htgo.hum.*
40: em.htgo.mus.*
41: em.htgo.other.*

Pred. No. is the number of results predicted by chance to have a

score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match %	Length	DB ID	Description
1	990	100.0	990	6	E17299
2	990	100.0	6881	1	D89064
3	990	100.0	44266	1	AB047548
4	990	100.0	199035	1	AB088420
5	988.4	99.8	6880	1	AB001723
6	871.6	88.0	20017	1	AB095952
7	130.8	13.2	8620	1	AF065891
8	117.2	11.8	5337	1	D83068
9	117.2	11.8	5337	6	E32813
10	109	11.0	4729	1	AF282898
11	109	11.0	4993	1	PF004052
12	107.8	10.9	5743	6	E39081
13	104.6	10.6	24448	1	AB024945
14	91.6	9.3	8154	1	AF157565
15	85	8.5	10580	1	AB102786
16	81.6	8.2	6658	1	AF001356
17	81.2	8.2	1026	1	FTS38289
18	73.8	7.5	4911	1	PST5663
19	73.8	7.5	4911	1	XSP012090
20	64.6	6.5	1180	1	PSETMOF
21	64.6	6.5	4729	6	AR362443
22	62.6	6.3	8456	1	RSP249207
23	56.6	5.7	7509	1	AF091716
24	54.6	5.5	11514	1	AF039533
25	54.6	5.5	12808	1	PSEORF1
26	53	5.4	6739	1	AY125981
27	53	5.4	9706	1	AF010471
28	53	5.4	83042	1	AF491307
29	53	5.4	101858	1	AY208917
30	50.2	5.1	1105	1	PAU86603
31	49.8	5.0	4355	1	FSENAPODNA
32	45.6	4.6	1058	1	AY048766
33	45.6	4.6	14462	1	AB004059
34	43	4.3	2000	6	AX655393
35	42.4	4.3	346510	1	AP003011
36	41	4.1	4799	1	AB098505
37	40.8	4.1	6652	1	AF360864
38	40.4	4.1	2000	6	AX655393
39	39.6	4.0	10506	1	AE008048
40	39.6	4.0	11109	1	AE009082
41	39.2	4.0	9783	1	ACA564846
42	39	3.9	4721	6	E32624
43	39	3.9	7823	1	AB024741
44	38.8	3.9	125020	9	AF429315
45	38.6	3.9	6298	1	AB006479

ALIGNMENTS

RESULT 1
E17299
LOCUS
DEFINITION
E17299
gDNA encoding oxidase which converts dibenzo-p-dioxin to a diphenyl ether triol.
ACCESION
E17299
VERSION
E17299.1 GI:5711982
KEYWORDS
JP 1998257895-A/8.
SOURCE
Pseudomonas sp.
ORGANISM
Bacteria; Proteobacteria.
REFERENCE
1 (bases 1 to 990)
AUTHORS
Omori, T. and Takami, K.
TITLE
OXIDASE GENE ORIGINATING FROM MICROORGANISM AND REMOVAL OF DIOXIN WITH THE SAME

E17299 990 bp DNA linear PAT 28-JUL-1999
gDNA encoding oxidase which converts dibenzo-p-dioxin to a diphenyl ether triol.

JOURNAL Patent: JP 1998257895-A 8 29-SEP-1998;
 ASahi CHEM IND CO., LTD
 OS Pseudomonas sp.
 PN JP 1998257895-A/8
 PD 29-SEP-1998
 PF 18-MAR-1997 JP 1997084401
 PI OMORI TOSHIO, TAKAMI KAZUTAKA
 PC C12N15/09, A62D3/00, B09C1/10, C02F3/34, C07H21/04, PC
 C12N1/21//C12N9/02,
 CC (C12N15/09, C12R1:38), (C12N1/21, C12R1:19), (C12N9/02, C12R1:19);
 CC strandedness: Double;
 CC topology: linear;
 CC hypothetical: No;
 CC anti-sense: No; Location/Qualifiers
 FH Key
 FT source 1..990
 FT /organism="Pseudomonas sp."
 FT /strain="CA10"
 FT Location/Qualifiers
 FT 1..990
 FT /organism="Pseudomonas sp."
 FT /mol_type="genomic DNA"
 FT /db_xref="taxon:306"

FEATURES
 source
 ORIGIN
 Query Match 100.0%; Score 990; DB 6; Length 990;
 Best Local Similarity 100.0%; Pred. No. 1.3e-256;
 Matches 990; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 1 ATGTACCAACTCAAAATTTGAAGGCGCAGCGCGCTGCGGCTCAGGGAAGAGCGCTG 60
 Db 1 ATGTACCAACTCAAAATTTGAAGGCGCAGCGCGCTGCGGCTCAGGGAAGAGCGCTG 60
 QY 61 TTGGTCTCAGCACTGCTTAATGGTATCGAATTCGATTCGATACGAGTGTGATCGGAGGTTC 120
 Db 61 TTGGTCTCAGCACTGCTTAATGGTATCGAATTCGATTCGATACGAGTGTGATCGGAGGTTC 120
 QY 121 CGAGTATCAAAATTCGAGTCTACTCGAAGGGAATGCCAATCAATGTGGCGGATGCTCCA 180
 Db 121 CGAGTATCAAAATTCGAGTCTACTCGAAGGGAATGCCAATCAATGTGGCGGATGCTCCA 180
 QY 181 GGACTTTCTCGGAGATCGTGAGAGGGCAACCGCATCTTGATCGCATGCCAGTGTGGCG 240
 Db 181 GGACTTTCTCGGAGATCGTGAGAGGGCAACCGCATCTTGATCGCATGCCAGTGTGGCG 240
 QY 241 CTCTCAGACCTCGGATCAAAAGTGCAGTGCAGGACAGTAGTACGTCCTCAAGCATTCCAATC 300
 Db 241 CTCTCAGACCTCGGATCAAAAGTGCAGTGCAGGACAGTAGTACGTCCTCAAGCATTCCAATC 300
 QY 301 TCAAGATGGAAGCGGAAGTTGTTGAGTCCGGCGCTTAATCAATGACCTGCTGTCCGTG 360
 Db 301 TCAAGATGGAAGCGGAAGTTGTTGAGTCCGGCGCTTAATCAATGACCTGCTGTCCGTG 360
 QY 361 CGATTACGCACTGATGGCGGACGAAATTTCTCCCGGCGGCTTCTGCTAGTAGAGGCA 420
 Db 361 CGATTACGCACTGATGGCGGCGGAAATTTCTCCCGGCGGCTTCTGCTAGTAGAGGCA 420
 QY 421 GAGCAGTTCGACGCGCTGTTGCGGCAATTCATGCGCAATTTAAAGAACCCCGAAGGC 480
 Db 421 GAGCAGTTCGACGCGCTGTTGCGGCAATTCATGCGCAATTTAAAGAACCCCGAAGGC 480
 QY 481 ATATGGAGTCTTATTAAGAGGATACCCAGAGCAATTTAGTCTTGGCTTTTCGAA 540
 Db 481 ATATGGAGTCTTATTAAGAGGATACCCAGAGCAATTTAGTCTTGGCTTTTCGAA 540
 QY 541 AATAGAAAGAGCGCTGCTCTATTTTGTAGCGGACCAATGGGCACATCTTCTTCGGT 600
 Db 541 AATAGAAAGAGCGCTGCTCTATTTTGTAGCGGACCAATGGGCACATCTTCTTCGGT 600
 QY 601 CCAGGACCGCGGCAAGAGTCTTTGATTCGGCGCGTGCAGGCTCTCGATGCGGCC 660
 Db 601 CCAGGACCGCGGCAAGAGTCTTTGATTCGGCGCGTGCAGGCTCTCGATGCGGCC 660

QY 661 GCTATTTCACGCGCCTCGATCGCGAAGACACAGACCGGTAAGTTGTTCTACGGCTCA 720
 Db 661 GCTATTTCACGCGCCTCGATCGCGAAGACACAGACCGGTAAGTTGTTCTACGGCTCA 720
 QY 721 AGAATTCGCGCGACGCTGTTGGTGGATCGATATCGATCGATCGATCGATCGATCGATCG 780
 Db 721 AGAATTCGCGCGACGCTGTTGGTGGATCGATATCGATCGATCGATCGATCGATCGATCG 780
 QY 781 GTCTGCCAGGCAAGTACGGAACACACGATAGCTTTCGCAAGGCGCCACCTGTTTATT 840
 Db 781 GTCTGCCAGGCAAGTACGGAACACACGATAGCTTTCGCAAGGCGCCACCTGTTTATT 840
 QY 841 CATCAGGTTGTCAGCGCGCTGTTGAAACCCCTACCGGAATACGAAATTTATCTTGC 900
 Db 841 CATCAGGTTGTCAGCGCGCTGTTGAAACCCCTACCGGAATACGAAATTTATCTTGC 900
 QY 901 GCTCACCGCTATGTCGACGCTACTCTCGTATGCTGCTCGGCAAGGCTGTTCCACGC 960
 Db 901 GCTCACCGCTATGTCGACGCTACTCTCGTATGCTGCTCGGCAAGGCTGTTCCACGC 960
 QY 961 GATCAAAATTCATTTTACGCAATTTTCTAA 990
 Db 961 GATCAAAATTCATTTTACGCAATTTTCTAA 990

RESULT 2
 D89064
 LOCUS
 DEFINITION Pseudomonas sp. gene for carbazole 1,9a-dioxygenase, meta-cleavage enzyme, meta-cleavage compound hydrolase, complete cds.
 ACCESSION D89064
 VERSION D89064.1 GI:2317677
 KEYWORDS reductase component of carbazole 1,9a-dioxygenase; ferredoxin component of carbazole 1,9a-dioxygenase; meta-cleavage compound hydrolase; meta-cleavage enzyme; terminal dioxygenase component of carbazole 1,9a-dioxygenase.
 SOURCE Pseudomonas sp.
 ORGANISM Pseudomonas sp.
 Bacteria; Proteobacteria.
 REFERENCE 1 (sites)
 AUTHORS Sato, S.I., Ouchiya, N., Kimura, T., Nojiri, H., Yamane, H. and Omori, T.
 TITLE Cloning of genes involved in carbazole degradation of Pseudomonas sp. strain CA10: nucleotide sequences of genes and characterization of meta-cleavage enzymes and hydrolase
 JOURNAL J. Bacteriol. 179 (15), 4841-4849 (1997)
 MEDLINE 97386424
 PUBMED 9244273
 REFERENCE 2 (sites)
 AUTHORS Sato, S.I., Nam, J.W., Kasuga, K., Nojiri, H., Yamane, H. and Omori, T.
 TITLE Identification and characterization of genes encoding carbazole 1,9a-dioxygenase in Pseudomonas sp. strain CA10
 JOURNAL J. Bacteriol. 179 (15), 4850-4858 (1997)
 MEDLINE 97386425
 PUBMED 9244274
 REFERENCE 3 (bases 1 to 6881)
 AUTHORS Omori, T.
 TITLE Direct Submission
 JOURNAL Submitted (14-NOV-1996) Toshio Omori, The University of Tokyo, Biotechnology Research Center; Yayoi 1-1-1, Bunkyo-ku, Tokyo 113, Japan [E-mail: assigyo@hongo.ecc.u-tokyo.ac.jp, Tel: 03-3812-2111 (ex.3067), Fax: 03-5802-3326]
 FEATURES
 source
 1..6881
 /organism="Pseudomonas sp."
 /mol_type="genomic DNA"
 /strain="CA10"
 /db_xref="taxon:306"
 572..1726
 /gene="carAa"
 572..1726
 /gene="carAa"

```
/codon_start=1
/transl_table=11
/product="terminal dioxygenase component of carbazole
1,9a-dioxygenase"
/protein_id="BAA21728.1"
/db_xref="GI:2317678"
/translation="WANYDEAILKRVKGNAPYVDALQFRNHWTFVMSKEIDGEPK
TLKLGELLNVRIDGKLYCLDRCLHRGVSQVKECTKSTTTCWHAWTYRWDG
VLCDILTPTSAIQIGKQKLTYPQOAEKGVFIYLGDDPPPLARDTPPNFLDDMEI
LGKNGI IKSNRWLA VENGDFDP SHIY IHKDSILVKNDLALPLGAPGDRKQOQTRVVD
DDVGRKGVYDLIGSHGVFVEGTIGGEVREGAYGEKIVANDISIMLPGLVKYNPPF
NPDMMQFSEWVPI DENTHY YFOTLGKPCANDEERKNYQEBESKWKWALGFGNDDI
WAREAMVDFYADDKGMVNEILLFEVDEAIVARKLASEHNQGIQTQAHVSG"
1838..2992
/gene="carAb"
CDS
1838..2992
/gene="carAb"
/codon_start=1
/transl_table=11
/product="terminal dioxygenase component of carbazole
1,9a-dioxygenase"
/protein_id="BAA21729.1"
/db_xref="GI:2317679"
/translation="WANYDEAILKRVKGNAPYVDALQFRNHWTFVMSKEIDGEPK
TLKLGELLNVRIDGKLYCLDRCLHRGVSQVKECTKSTTTCWHAWTYRWDG
VLCDILTPTSAIQIGKQKLTYPQOAEKGVFIYLGDDPPPLARDTPPNFLDDMEI
LGKNGI IKSNRWLA VENGDFDP SHIY IHKDSILVKNDLALPLGAPGDRKQOQTRVVD
DDVGRKGVYDLIGSHGVFVEGTIGGEVREGAYGEKIVANDISIMLPGLVKYNPPF
NPDMMQFSEWVPI DENTHY YFOTLGKPCANDEERKNYQEBESKWKWALGFGNDDI
WAREAMVDFYADDKGMVNEILLFEVDEAIVARKLASEHNQGIQTQAHVSG"
3050..3322
/gene="carBa"
CDS
3050..3322
/gene="carBa"
/codon_start=1
/transl_table=11
/product="subunit of meta-cleavage enzyme"
/protein_id="BAA21730.1"
/db_xref="GI:2317680"
/translation="MARVEVDLLIQDMSKEGLIGRVIDTPSDVFEVYGLTPPERTAL
LEGTPQALASIGVHPILQMHVLMYKNPEMATHVSIKDYSDMLKGA"
3322..4131
/gene="carBb"
CDS
3322..4131
/gene="carBb"
/codon_start=1
/transl_table=11
/product="catalytic subunit of meta-cleavage enzyme"
/protein_id="BAA21731.1"
/db_xref="GI:2317681"
/translation="MGKIVAAGCTSHILMSPKGCEBSAARVVNGIABLRRLKEARPD
VLVITSDHMFNLSMOPRFVVGIAISYTPMGDMIDPRDLVPGSRREVGRALQADE
DGFLLCOAEYSLDHGINIPILFMGKEIPVVPVINIDPIPSARRCVAALESIRQ
AIEKRTPDGCRVAVVAGGLSHWLCVPRHGEVSEKFDHVMVDELVRGNAEKLVMGNE
AIIIDGGNAGVEILLTWIMAAVASBASSEKVFYEAMTQWFTIGGMBFHVK"
5014..5337
/gene="carAc"
CDS
5014..5337
/gene="carAc"
/codon_start=1
/transl_table=11
/product="ferredoxin component of carbazole
1,9a-dioxygenase"
/protein_id="BAA21733.1"
/db_xref="GI:2317683"
/translation="MNQIWLKVCASDMQPGTIRRVNRVGAAPLAVRVGDQFVATED
TCTHGIASLSEGTLDGDIIECPFHGGAFNVCTGMPASSPCTVPLGVFEVEVKEGEVYV
AGEKK"
5718..6707
/gene="carAd"
CDS
5718..6707
/gene="carAd"
/codon_start=1
```

```
/transl_table=11
/product="ferredoxin reductase component of carbazole
1,9a-dioxygenase"
/protein_id="BAA21735.1"
/db_xref="GI:2317685"
/translation="MYOLKIEGQAPCTCGSKSLVALANGIFPFYECASGCGVCK
FELEGNVQSMWPDAPGLSRDRKGRHLACQVALSDLRIRKVAQDKVPTIPISR
MEAEVVEVRLATHDLLSVRLRTDGPANFLPGQFCLVEAQLPGVVRAYSMANLKNPEG
IMEFYIKRVPYTRGRFSPWLPENRKEGARLFLTGMPTSFPRGTRKSLCIGGAGLSY
AAAIARASMETDKPKLVFGSRTPRDAVRWIDIDIDEDKLEVVQAVTETDLSLWQGP
TGFIHQVWDAALETILPEVEIYLAGEPPWVDATVRMLLKGKVPDRDQIHDAFF"
ORIGIN
Query Match 100.0%; Score 990; DB 1; Length 6881;
Best Local Similarity 100.0%; Pred. No. 1.5e-256;
Matches 990; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 ATGTACCAACTCAAAATTGAAGGGCAAGCGCCAGGACCTGCGGCTCAGGGAAGAGCCTG 60
Db 5718 ATGTACCAACTCAAAATTGAAGGGCAAGCGCCAGGACCTGCGGCTCAGGGAAGAGCCTG 5777
QY 61 TTGGTCTCAGCACTTGTCTAATGCTATCGGATTTCCGTACGAGTGTGCATCGGAGGTTGC 120
Db 5778 TTGGTCTCAGCACTTGTCTAATGCTATCGGATTTCCGTACGAGTGTGCATCGGAGGTTGC 5937
QY 121 GGAGTATGCAAAATTCGAGTACTCGAAGGGAATGTCCAATCAATGATGGCGGATGCTCCA 180
Db 5938 GGAGTATGCAAAATTCGAGTACTCGAAGGGAATGTCCAATCAATGATGGCGGATGCTCCA 5997
QY 181 GGACTTTCTTCGCGAGATCGTGAGAGGGCAACCGCCATCTTCGATGCCAGTCCGTTGCG 240
Db 5898 GGACTTTCTTCGCGAGATCGTGAGAGGGCAACCGCCATCTTCGATGCCAGTCCGTTGCG 5957
QY 241 CTCTCAGACTCGCGGATCAAGTTCGAGTCCGCGGCTAACTCATCACCTGCTGTCGCTG 300
Db 5958 CTCTCAGACTCGCGGATCAAGTTCGAGTCCGCGGCTAACTCATCACCTGCTGTCGCTG 6017
QY 301 TCAAGAAATGGAAGCGGAAGTTGTGAGTCCGCGGCTAACTCATCACCTGCTGTCGCTG 360
Db 6018 TCAAGAAATGGAAGCGGAAGTTGTGAGTCCGCGGCTAACTCATCACCTGCTGTCGCTG 6077
QY 361 CGATTACGCACTGATCGGCGAGCAAAATTCCTCCCGGCGAGTTCCTCTAGTAGAGGCA 420
Db 6078 CGATTACGCACTGATCGGCGAGCAAAATTCCTCCCGGCGAGTTCCTCTAGTAGAGGCA 6137
QY 421 GAGCAGTTGCGAGCGCTGTTGCGGATATTCAATGGCGAATTTAAGAACCCCGAAGGC 480
Db 6138 GAGCAGTTGCGAGCGCTGTTGCGGATATTCAATGGCGAATTTAAGAACCCCGAAGGC 6197
QY 481 ATATGGAGTTCTPATTTAAGAGGTPACCCACAGACGATTTAGTCTCTTGGCTTTTCGAA 540
Db 6198 ATATGGAGTTCTPATTTAAGAGGTPACCCACAGACGATTTAGTCTCTTGGCTTTTCGAA 6257
QY 541 AATAGAAAAGAGCGCTCTCTATTTTTCGCGGACCAATGGGACATCTTCTTCCGT 600
Db 6258 AATAGAAAAGAGCGCTCTCTATTTTTCGCGGACCAATGGGACATCTTCTTCCGT 6317
QY 601 CCAGGACCCGGCGAAGAGTCTTTTCATTTGGCGCGGTGCGCGGCTCTCTGATGCGGC 660
Db 6318 CCAGGACCCGGCGAAGAGTCTTTTCATTTGGCGCGGTGCGCGGCTCTCTGATGCGGC 6377
QY 661 GCTATTGCGCGCTCTGATGCGGCAACAGACAGCCCGGTAAAGTTGTTCTACGGCTCA 720
Db 6378 GCTATTGCGCGCTCTGATGCGGCAACAGACAGCCCGGTAAAGTTGTTCTACGGCTCA 6437
QY 721 AGAATCTCCGCGACGCTGTTCCGTGATTCGATATCGACATCGATGAGGACAAAGCTTGAG 780
Db 6438 AGAATCTCCGCGACGCTGTTCCGTGATTCGATATCGACATCGATGAGGACAAAGCTTGAG 6497
QY 781 GTCGTCCAGGAGTTACGGAAGACAGGATAGCTTTTGGCAAGGGCCCACTGGTTTTATT 840
Db 6498 GTCGTCCAGGAGTTACGGAAGACAGGATAGCTTTTGGCAAGGGCCCACTGGTTTTATT 6557
```

QY 841 CATCAGGTTGTCGACGCGAGCGTCTGTTGAACCCCTACCGGAATACGAAATTTATCTTGC 900
Db 6558 CATCAGGTTGTCGACGCGAGCGTCTGTTGAACCCCTACCGGAATACGAAATTTATCTTGC 6617
QY 901 GGTCCACCGCTATGTCGAGCGTACTGTCGATGCTGCTCGCAAGGTTCTCCACGC 960
Db 6618 GGTCCACCGCTATGTCGAGCGTACTGTCGATGCTGCTCGCAAGGTTCTCCACGC 6677
QY 961 GATCAAAATTCATTTTGACGCAATTTTCTAA 990
Db 6678 GATCAAAATTCATTTTGACGCAATTTTCTAA 6707

RESULT 3
AB047548 44266 bp DNA linear BCT 03-MAY-2002
LOCUS Pseudomonas resinovorans strain CA10 DNA, car gene cluster flanking
DEFINITION region.
ACCESSION AB047548
VERSION AB047548.1 GI:13094152
KEYWORDS
SOURCE Pseudomonas resinovorans
ORGANISM Pseudomonas resinovorans
Bacteria; Proteobacteria; Gammaproteobacteria; Pseudomonadales;
Pseudomonadaceae; Pseudomonas.
REFERENCE
1 Sato,S.I., Ouchiyaana,N., Kimura,T., Nojiri,H., Yamane,H. and
Omori,T.
Cloning of genes involved in carbazole degradation of Pseudomonas
sp. strain CA10: nucleotide sequences of genes and characterization
of meta-cleavage enzymes and hydrolase
J. Bacteriol. 179 (15), 4841-4849 (1997)
97386424
PUBMED 9244273
REFERENCE
2 Sato,S.I., Nam,J.W., Kasuga,K., Nojiri,H., Yamane,H. and Omori,T.
Identification and characterization of genes encoding carbazole
1,9a-dioxygenase in Pseudomonas sp. strain CA10
J. Bacteriol. 179 (15), 4850-4858 (1997)
97386425
PUBMED 9244274
REFERENCE
3 Nojiri,H., Sekiguchi,H., Maeda,K., Urata,M., Nakai,S., Yoshida,T.,
Habe,H. and Omori,T.
Genetic characterization and evolutionary implications of a car
gene cluster in the carbazole degrader Pseudomonas sp. strain CA10
J. Bacteriol. 183 (12), 3663-3679 (2001)
21264379
PUBMED 11371531
REFERENCE
4 (bases 1 to 44266)
Nojiri,H., Omori,T. and Habe,H.
Direct Submission
Submitted (21-AUG-2000) Hideaki Nojiri, The University of Tokyo,
Biotechnology Research Center; 1-1-1 Yayoi, Bunkyo-ku, Tokyo
113-8657, Japan (E-mail:anojiri@mail.ecc.u-tokyo.ac.jp,
Tel:81-3-5841-3064 Fax:81-3-5841-8030)
Location/Qualifiers
1. .44266
/organism="Pseudomonas resinovorans"
/mol_type="genomic DNA"
/strain="CA10"
/db_xref="taxon:53412"
complement(214. .1929)
/gene="ORF31"
complement(214. .1929)
/gene="ORF31"
/function="putative membrane spanning protein"
/codon_start=1
/transl_table=11
/product="putative ABC transporter subunit"
/protein_id="BAB32742.1"
/db_xref="GI:13094153"
/translation="MGVINLAHGAFFYMWGAFRCPTPRPERLVLARRVGGSSARPVTR

gene

CDS

complement(2085. .3245)

/gene="ORF30"

complement(2085. .3245)

/gene="ORF30"

/function="putative ABC transporter subunit"

protein"

/codon_start=1

/transl_table=11

/product="putative ABC transporter subunit"

/protein_id="BAB32743.1"

/db_xref="GI:13094154"

/translation="MKKQLVGTLLAPVANPALRPAQDPVRIPTLSTPAGVLGD

SRDFOALIDQEGKLGKGVFVSLVEDDGLQARAKQIIDMSLDGSLYGVIFSNV

LAAYVNTATKDFYISNTPSSLAGCKPQFVASYQNDTHMAGVANDLGGYK

WVILAPYQGRDALGKRTFKGEVETITKLNQDLDFVELARSLADPAIDFQFHP

GGAGINFAKQYSSGLNKSIPIWVVFVSMDBERMLAATGDVAKGMNITLWNPSTNPA

NOAFVKANDKVPVTVAAQTVTARLIGAAKTVQDGNLKDPEALRSALRNVTFDS

IRGPAFGKQHAVIDWYLLRVEADANGKLVQVPVQVIAQAQVDSFAECL"

complement(3405. .4385)

/gene="tnpA3"

complement(3405. .4385)

/gene="tnpA3"

/note="TnpA3"

/codon_start=1

/transl_table=11

/product="transposase"

/protein_id="BAB32744.1"

/db_xref="GI:13094155"

/translation="WKQMTFADARYAGKQKXELFLIEMDRVVPWKGLIALIEPHY

PKGGRGPAYPLMAMLRVHLQWVFGYSDPAMEALYETTLRQFAGINLERIDPETT

ILNFRRLLEKELAAIGILAVINGVLGRGLSROGTIVDTLINAPSTKNGDKRDP

EMHCTKGNQVYFCMKAHIGADDEGLVHVSVGTAAVADTVQDKLLHGVENVVAD

AGYGVKRPHEGSEVWQVAAARSTYKLDKRSVLYKAKKIEKAKAQVRAKVEHS

FRVIRQGYTKVRFRLGKLTAVLTFALSNLWMAHRLITNAGEVRL"

complement(4543. .5571)

/gene="antC"

complement(4543. .5571)

/gene="antC"

/note="AntC"

/codon_start=1

/transl_table=11

gene

CDS

complement(5590. .6081)

/gene="antB"

complement(5590. .6081)

/gene="antB"

/note="AntB"

/codon_start=1

/transl_table=11

/product="small subunit of terminal oxygenase component of

anthranilate 1,2-dioxygenase"

/protein_id="BAB32746.1"

/db_xref="GI:13094157"

/translation="GI:13094157"

gene

CDS

gene

CDS

complement(5590. .6081)

/gene="antB"

complement(5590. .6081)

/gene="antB"

/note="AntB"

/codon_start=1

/transl_table=11

/product="small subunit of terminal oxygenase component of

anthranilate 1,2-dioxygenase"

/protein_id="BAB32746.1"

/db_xref="GI:13094157"

FEATURES
source


```
/translation="MNSQLQYRVQFLYRNAECDAQWDYTYLSPFSDSDSEFLHPWD
SEHYTPDKPRAMSLIYVNPGRGLEDVRFIRITGKSASTVPMPTRIHLHLSNVRIAPLA
DELEVQVNHWTLYRILQTABQFFGRATYRLPDGESWKARKHVLLNDNTINSVLDF
YHL"
gene
CDS
complement (6084. .7487)
/genes="antA"
complement (6084. .7487)
/genes="antA"
/notes="antA"
/codon_start=1
/transl_table=11
/product="large subunit of terminal oxygenase component of
anthranilate 1,2-dioxygenase"
/protein_id="BAB32747.1"
/db_xref="GI:13094156"
/translation="MSSARSVEQWNFIIEGCLDRPADGVFRIARDIFTPDPLDLEM
ELIFEKWIYACHSESTANKDFMTFRAGQPMIITRDGEGQLNALINACQHRGTTLT
RVSGNOSTFTCPFHAWCYKSDGLVVKAPFEGFEGFKATRGKLRKARIESYKGFYF
ISLDVGDSGLFEGDAKVFDMVAQSPTEGELEVLPGKSAYTYDGNWKLQENGLD
GYHSTVHYVAVTQHQVNSEKGLTSADTLDYSKLKAGDKETDQGWFAFNHGSV
LFSMPNPSVPQVATIMPRLLQVYGOGKQMMHRLRNLIYPSMFELDQISQLEI
IRPIANWTEINSCLGVGSDADRENRIQFEDPFNVSMGTPDDLVHFRZARQRF
CARLERMSDISRGVKWVEGPTANSETIGISPVLUTGEFTHEGLYVNHQGNWOFILLE
GLARAEQEQPLKLRV"
gene
CDS
complement (7829. .8809)
/genes="tnpA2"
complement (7829. .8809)
/genes="tnpA2"
/notes="tnpA2"
/codon_start=1
/transl_table=11
/product="transposase"
/protein_id="BAB32748.1"
/db_xref="GI:13094159"
/translation="MKQMTFAAEVAGKQKOTRKEFLIEMDRVVPNKGILALIEPHY
PKGEGRPAYPLMAMLRVHLNQWHPGYSDDPAMEALYETTLRQFAGLSLERIDETIT
LHNPRLKEHLARGILAVINGLDRGLSLRGQTIVDATLINAPSSTKNRQGRDP
EMHQKGNQNYFGMAHIGVDDSGLVSHVSTAANVADVTDVKLLHGDENWVGAD
AYTGVRKPEHREGVETWQAAARSTYKLDKRSALYKAARKTEKAKQVRAKVEHP
FRVIRQFYGVKVRGLAKNTAQLVTLFALSNLMMARRHLLANAGEVRL"
9062. .10318
/genes="ORF24"
9062. .10318
/genes="ORF24"
/function="channel-forming protein"
/codon_start=1
/transl_table=11
/product="phak-like protein"
/protein_id="BAB32749.1"
/db_xref="GI:13094160"
/translation="MFQASWFRGMQTASLALMTPMLAQADFVEDSOLALGLRNFYI
DRFKGDMPNSRVGSWTQGFDRFNSGTETGLQFALDASMQAYRLDGGGGRGPD
IIPSTRQEQATDYRGQYTAKVRYSKTELKYGHEFTLPVAYTDDSRQLVITYHGT
LIESKEVDKLTITGGRFTEISGRSSSEKMYLFNGPDVQRSDGLNFGGASVAFTPA
LGTGYFGGLEDIYRQHYLGATYLDLGLSGVALKIDRLYFNNKEDGALYGRINDRSY
GAMTTLRGAHSGFLGYQRMGLGDSFTPLNGVAPQPYLVNNSAVAFVKNPKSQVARY
DYDFAGLGLPGRLMTRMYRSGSDIDEGALSDNSESERNIYLSYVVQEGPLKNLQVELR
NINWFRHNGDFENRLITTYTWKFW"
10639. .11655
/genes="ORF23"
10639. .11655
/genes="ORF23"
/codon_start=1
/transl_table=11
/product="putative regulatory protein"
/protein_id="BAB32750.1"
/db_xref="GI:13094161"
/translation="MMSTSLDVHVRDRIERYDLNGARQWMAGICGPHLLKTYQPNRI
QPHHSANVLKSMSTILGVIEYGTDTVVGIEDVHLSYSLSLPLVGEQHLKSGDQCVT
```

Query Match 100.0%; Score 990; DB 1; Length 44266;
Best Local Similarity 100.0%; Pred. No. 1.6e-256; Mismatches 0; Indels 0; Gaps 0;
Matches 990; Conservative 0;

```
QY 1 ATGTACCAACTCAAAATTGAAGGGCAAGCGCCAGGGACCTCGCGCTCAGGGGAAGACCTG 60
DB 33655 ATGTACCAACTCAAAATTGAAGGGCAAGCGCCAGGGACCTCGCGCTCAGGGGAAGACCTG 33714
QY 61 TTGGTCTCAGCACTTGTCTAATCGTATTCGGATTTTCGTACAGTGTGCATCGCGAGGTTCG 120
DB 33715 TTGGTCTCAGCACTTGTCTAATCGTATTCGGATTTTCGTACAGTGTGCATCGCGAGGTTCG 33774
QY 121 GGAGTATGCAAAATTCGAGTTACTCGAAGGGAATGTCCAATCAATGTGGCCGAGTGTCTCA 180
DB 33775 GGAGTATGCAAAATTCGAGTTACTCGAAGGGAATGTCCAATCAATGTGGCCGAGTGTCTCA 33834
QY 181 GGACTTTCTTCGGGAGATCGTCGAGAGGCAACCGCATCTTTCGATGCGCAGTGTCTGTGG 240
DB 33835 GGACTTTCTTCGGGAGATCGTCGAGAGGCAACCGCATCTTTCGATGCGCAGTGTCTGTGG 33894
QY 241 CTCTCAGACTCGCGATCAAAATTCGAGTTCGAGGCAACAGTACGTCCCAACGATTCCAATC 300
DB 33895 CTCTCAGACTCGCGATCAAAATTCGAGTTCGAGGCAACAGTACGTCCCAACGATTCCAATC 33954
QY 301 TCAGAAATGGAAGCGGCAAGTTCGTAGGTTCGGGCGGTAACTCATGACCTGTCTCCG 360
DB 33955 TCAGAAATGGAAGCGGCAAGTTCGTAGGTTCGGGCGGTAACTCATGACCTGTCTCCG 34014
QY 361 CGATTACGCACTGATGGCGCAGCAAAATTTCTCCCGCGCAGTTCTGCTAGTAGAGGCA 420
DB 34015 CGATTACGCACTGATGGCGCAGCAAAATTTCTCCCGCGCAGTTCTGCTAGTAGAGGCA 34074
QY 421 GACGAGTTCGAGCGCTGTTCGCGCATATTCAATGGCGCAATTTAAAGAACCCCGAAGGC 480
DB 34075 GACGAGTTCGAGCGCTGTTCGCGCATATTCAATGGCGCAATTTAAAGAACCCCGAAGGC 34134
QY 481 ATATGGAGTTCATATTAAGAGGGTACCCACAGAGGAGTATTTAGTCTCTTGGCTTTTCGAA 540
DB 34135 ATATGGAGTTCATATTAAGAGGGTACCCACAGAGGAGTATTTAGTCTCTTGGCTTTTCGAA 34194
QY 541 AATAGAAAAGAGCGCGCTGTCTATTTTGAACGGGACCAATGGGACACATCTTCTTCGCT 600
DB 34195 AATAGAAAAGAGCGCGCTGTCTATTTTGAACGGGACCAATGGGACACATCTTCTTCGCT 34254
QY 601 CCAGGACCGCGCGCAAGAGTCTTTGCATTGGCGCGGTGCGCGGCTCTCGTATGCGGCC 660
DB 34255 CCAGGACCGCGCGCAAGAGTCTTTGCATTGGCGCGGTGCGCGGCTCTCGTATGCGGCC 34314
QY 661 GCTATTGACGCGCTCGATGCGCGCAACAGACACCGCGTAAAGTTGTTCTACGGCTCA 720
DB 34315 GCTATTGACGCGCTCGATGCGCGCAACAGACACCGCGTAAAGTTGTTCTACGGCTCA 34374
QY 721 AGAATCCGCGCGCAGCTGTTTCGTGATCGATATCGATCGATGAGGACCAAGCTTGAG 780
DB 34375 AGAATCCGCGCGCAGCTGTTTCGTGATCGATATCGATCGATGAGGACCAAGCTTGAG 34434
QY 781 GTCGTCCAGCAGTTACGGGAACACGAGTACGCTTTGGCAAGGCGCCACTGTTTTATT 840
DB 34435 GTCGTCCAGCAGTTACGGGAACACGAGTACGCTTTGGCAAGGCGCCACTGTTTTATT 34494
QY 841 CATCAGTTCGCGACGCGCTGTTGAAACCTCCGGAATACGAAATTTATCTTGCC 900
DB 34495 CATCAGTTCGCGACGCGCTGTTGAAACCTCCGGAATACGAAATTTATCTTGCC 34554
QY 901 GGTCACCGCTATGTCGACGCTACTGTCCGTATGCTGTCTCGCAAGGGTGTTCACGC 960
DB 34555 GGTCACCGCTATGTCGACGCTACTGTCCGTATGCTGTCTCGCAAGGGTGTTCACGC 34614
QY 961 GATCAAAATTCATTTGACGCAATTTTCTAA 990
DB 34615 GATCAAAATTCATTTGACGCAATTTTCTAA 34644
```

RESULT 4
AB088420
LOCUS AB088420 199035 bp DNA circular BCT 03-MAY-2003

DEFINITION Pseudomonas resinovorans plasmid pCARI, complete sequence.
 ACCESSION AB088420.1 GI:26106080
 VERSION AB088420.1
 KEYWORDS Pseudomonas resinovorans
 SOURCE Pseudomonas resinovorans
 ORGANISM Bacteria; Proteobacteria; Gammaproteobacteria; Pseudomonadales; Pseudomonadaceae; Pseudomonas.

REFERENCE
 1. Maeda, K., Nojiri, H., Shintani, M., Yoshida, T., Habe, H. and Omori, T. Complete nucleotide sequence of carboxyle/dioxin-degrading plasmid pCARI in Pseudomonas resinovorans strain CAL0 indicates its mosaicity and the presence of large catabolic transposon Tn4676 J. Mol. Biol. 326 (1), 21-33 (2003)
 MEDLINE 22435905
 PUBMED 12547188

REFERENCE
 2. (bases 1 to 199035)
 AUTHORS Nojiri, H., Shintani, M. and Maeda, K.
 TITLE Direct Submission
 JOURNAL Submitted (22-JUL-2002) Hideaki Nojiri, Biotechnology Research Center, The University of Tokyo; 1-1-1 Yayoi, Bunkyo-ku, Tokyo 113-8657, Japan (E-mail:anojiri@mail.ecc.u-tokyo.ac.jp, Tel:81-3-5841-3064, Fax:81-3-5841-8030)
 FEATURES Location/Qualifiers

1. 199035
 /organism="Pseudomonas resinovorans"
 /mol_type="genomic DNA"
 /db_xref="taxon:53412"
 /plasmid="pCARI"
 join(185744..199035,1..59548)
 /transposon="Tn4676"
 complement(join(198367..199035,1..81))
 /gene="ORF18"
 complement(join(198367..199035,1..81))
 /gene="ORF188"
 /function="putative ABC-transporter subunit"
 /codon_start=1
 /evidence=not experimental
 /transl_table=11
 /product="putative ABC-transporter subunit"
 /protein_id="BAC41710.1"
 /db_xref="GI:26106270"

repeat_region
 gene
 CDS
 /translation="MNPQLSVALEKTFGAVKATNGASFDIAPGELHAIIGPNAGKAS
 TLAQAGEIRPKGRIFLNGQDITVPAHQRLGLARAFQVSQLYPFTLENNVAV
 AIAARDQOTGLKWPASDOSLLAPADYILARVGLRGHAQVSALSGBERGLEIAL
 ALAQAESVLLIDEPMAQGAQESARMTQLLELKGRYAILLVEHMDMAVFAALADRTY
 LVVGQTLTGSQVRNDPHVRAAYLGEHA"
 complement(1108..1929)
 /gene="ORF31"
 complement(1108..1929)
 /gene="ORF31"
 /function="putative membrane spanning subunit"
 /codon_start=1
 /evidence=not experimental
 /transl_table=11
 /product="putative ABC transporter subunit"
 /protein_id="BAC41521.1"
 /db_xref="GI:26106081"

gene
 CDS
 /translation="MGVINLAHGAFFVMGAFCAAYTASLSGSFWLAGLAAGVLSALYGL
 LLTETSMLRYNRDLDQVLAFLAVVFTNELITVLFGSSPPMPTPEWYSFVEVL
 PGLMVPVSRLLFTAGAAVALAMWLLNHLRLGLIPAGADHMGVAGLVNRYLYT
 LVFVGCVLGLAGMAAPLLSVEIGKGEKVLITTFVIVVGVGVGRGLAGALLIG
 MVDGLRAYLPQLLDQLLPAREVSGTSLGGLISASAYIVMAVILLVPRGLLPARA"
 complement(2084..3244)
 /gene="ORF30"
 complement(2084..3244)
 /gene="ORF30"
 /evidence=not experimental
 /transl_table=11
 /product="putative ABC transporter subunit"
 /protein_id="BAC41522.1"
 /db_xref="GI:26106082"

/translation="MKKQVGTLLAASCLASLAQAQDPVIRIGFITTLSTPAGVLGSD
 SRGFOAIIDQGGKLGVPVSLIVDDGLQARAKQIDMSLQDGLSLVYGVPSNV
 LAVVNTATKDGFIYISNTGSPSLAGKQCPHYFVASYQNDTHEMAGVAANDLGYKK
 MVLLAPNYQGRDALSGFKRTFKGEVTEIYTKNLQDFSVELARVSLADPAIFQFHP
 GGAGINFAKQYSSGLNKSIPMVVFSMDERMLAATGDAKGNVITLWNPSTNPA
 NQAFKADNDKYKRVPTVYAAQTYDARLIGAAALKTVGNLKDPEALRSALRNTFDS
 IRGFAFGKQHAVIDWYLLRVEADANGKLVQVPTQITAKAQVDSFAAECAL"
 3255..3258
 /rpt_type=direct
 /transl_table=11
 /product="transposase of ISPre2"
 /protein_id="BAC41523.1"
 /db_xref="GI:26106083"

repeat_unit
 repeat_region
 repeat_unit
 gene
 CDS
 /translation="MKOMTFADABYAGKQKELFLIEMDRVVPWKGLIALTEPHY
 PKGGSPAYFLMAMLVHLLQWFGYSDPAMEALYETTLIQFACGLNLERIDETT
 ILNFRLLKHELAAAGILAVINGVLGRGLSROGTIVDAILINAPSTKXKDKRDP
 EMHQTKNQIYFGKAHIGADDESGLVHVSVVGPAANVADVTDVKLHLGDNVVCAD
 AGYTGKRGPEHEGRIWQVARRSTYKLDKESVLYKARKIEKAKQAQVRAKVEHP
 FRVIRQFGYTKVRFGRLAKNTAQLVTLFALSNNLMARRHLLTNAGEVRL"
 4433..4448
 /rpt_type=inverted
 /transl_table=11
 /product="reductase component of anthranilate"
 /protein_id="BAC41524.1"
 /db_xref="GI:26106084"

gene
 CDS
 /translation="MSHKVAFSADGKTLFFPVQSHETLLDAALRNGINIPLDREGV
 CQTCGRCSQYQVYDDEALSAEDLRQRMKLTCTQTRVSSDAAFDFPSSACNAP
 GPKGLKTVRAVEQVSNQTLHLHDAGADGQGLDFLQGVARLQIPGTHGORSYFAN
 RPSGNOLQFLIRLLPGLGMSNVIRESLIGDELLLEAPLGAFLRVHVDKPLVFAVG
 TGLSAPFLMLDEIAERGCGHPHYLVGVNRNAOLCETQRIAAVAERLPGFRFTPLV
 DPSPDWGHGRGIIITEHFLAEFRQALDMYLCGPPMVESIKQWEDTQALDQTLIYE
 KFTQSNIT"
 complement(5587..6078)
 /gene="antB"
 complement(5587..6078)
 /gene="antB"
 /note="antB"
 /codon_start=1
 /transl_table=11
 /product="small subunit of terminal oxygenase component of
 anthranilate 1,2-dioxygenase"
 /protein_id="BAC41525.1"
 /db_xref="GI:26106085"

gene
 CDS
 /translation="MNSQLOXYRVEQFLYRNAEACDAQDMDTVLSWFSDDSEHLPMQD
 SEHETIRPKAMSLIYFPNGGLEDNRVFRIRTKSASTVMPRTILLHSNRIAPLA
 DDELEQVNWHTLYRLQTAQFFGRATYRURPDGSEWIKARKHLLINDTINSVLDF
 YHL"
 complement(6081..7484)
 /gene="antA"
 complement(6081..7484)
 /gene="antA"
 /note="antA"
 /codon_start=1
 /transl_table=11
 /product="large subunit of terminal oxygenase component of

Query Match	100.0%;	Score 980;	DB 1;	Length 199035;
Best Local Similarity	100.0%;	Pred. No. 1.7e-256;		
Matches 990;	Conservative 0;	Mismatches 0;	Indels 0;	Gaps 0;
1	ATGTACCAACTCAAAATTGAAGGCGAAGCCGACGACCTGGGCTCAGGGAAGACCTG	50		
33633	ATGTACCAACTCAAAATTGAAGGCGAAGCCGACGACCTGGGCTCAGGGAAGACCTG	33692		
61	TTGGTCTCAGCACTTGCTAAATGGTATCGGATTTCCGTACGAGTGTGCATCGGAGGTTGC	120		
33693	TTGGTCTCAGCACTTGCTAAATGGTATCGGATTTCCGTACGAGTGTGCATCGGAGGTTGC	33752		
121	GGAGTATGCAAAATTCGAGTTTACTCGAAGGGAATGCCAATCAATGTGGCCGATGCTCCA	180		
33753	GGAGTATGCAAAATTCGAGTTTACTCGAAGGGAATGCCAATCAATGTGGCCGATGCTCCA	33812		
181	GGACTTTCTTCGCGAGATCGTGAGAAGGCCAACCCGCATCTTGCATGCCAGTGCCTTGGC	240		
33813	GGACTTTCTTCGCGAGATCGTGAGAAGGCCAACCCGCATCTTGCATGCCAGTGCCTTGGC	33872		

6257	AATGAAAGAGGCGCTGCTATTTTTCACGGGACCAATGGGCACATCTTTCTCCGT	6316
601	CCAGGACCGCGCGAAGAGCTTTTGATTCGGCGCGGTGCGGCTCTCGTATGCGGCC	660
6317	CCAGGACCGCGCGAAGAGCTTTTGATTCGGCGCGGTGCGGCTCTCGTATGCGGCC	6376
661	GCTATTGACGCGCTCGATCGCGGAAACAGACAGCCGGTAAAGTTGTTCTACGGCTCA	720
6377	GCTATTGACGCGCTCGATCGCGGAAACAGACAGCCGGTAAAGTTGTTCTACGGCTCA	6436
721	AGAACTCCGCGGACGCTGTTTCGGTGGATCGATATCGATCGATGAGCAACGCTTGAG	780
6437	AGAACTCCGCGGACGCTGTTTCGGTGGATCGATATCGATCGATGAGCAACGCTTGAG	6496
781	GTCTGTCAGGCGATACCGAAGACACGATAGCTTTGCAAGCGCCACCTGGTTTATT	840
6497	GTCTGTCAGGCGATATCGAAGACACGATAGCTTTGCAAGCGCCACCTGGTTTATT	6556
841	CATCAGGTTGTCGACGCGCTGTTGAAACCTTACCGAATACGAAATTAATCTTGCC	900
6557	CATCAGGTTGTCGACGCGCTGTTGAAACCTTACCGAATACGAAATTAATCTTGCC	6616
901	GGTCACCGCTATGTCAGCGTACTGTCGGTATGCTGCTCGCAAGGGTGTTCACGC	960
6617	GGTCACCGCTATGTCAGCGTACTGTCGGTATGCTGCTCGCAAGGGTGTTCACGC	6676
961	GATCAAAATCAATTTGACGCATTTTCTAA	990
6677	GATCAAAATCAATTTGACGCATTTTCTAA	6706
RESULT 6		
AB095952		
LOCUS	AB095952 20017 bp DNA linear BCT 04-FEB-2003	
DEFINITION	Janthinobacterium sp. J3 car operons, partial and complete cds.	
ACCESSION	AB095952	
VERSION	AB095952.1 GI:28201193	
KEYWORDS	Janthinobacterium sp. J3	
SOURCE	Janthinobacterium sp. J3	
ORGANISM	Bacteria; Proteobacteria; Betaproteobacteria; Burkholderiales; Oxalobacteraceae; Janthinobacterium.	
REFERENCE	1	
AUTHORS	Widada, J., Nojiri, H., Nakai, S., Inoue, K., Shintani, M., Yoshida, T., Habe, H. and Omori, T.	
TITLE	Molecular diversity of carbazole-utilizing bacteria and their catabolic genes	
JOURNAL	Unpublished	
REFERENCE	2	
AUTHORS	Inoue, K., Nojiri, H., Nakai, S., Endoh, T., Urata, M., Ashikawa, Y., Saito, Y., Yoshida, T., Habe, H. and Omori, T.	
TITLE	Divergent structures of carbazole-degrading car operons isolated from several gram-negative bacteria	
JOURNAL	Unpublished	
REFERENCE	3 (bases 1 to 20017)	
AUTHORS	Nojiri, H. and Inoue, K.	
TITLE	Direct Submission	
JOURNAL	Submitted (11-NOV-2002) Hideaki Nojiri, Biotechnology Research Center, The University of Tokyo; 1-1-1, Yayoi, Bunkyo-ku, Tokyo 113-8657, Japan (E-mail:anojiri@mail.ecc.u-tokyo.ac.jp, Tel:81-3-5841-3067, Fax:81-3-5841-8030)	
FEATURES		
source		
1. .20017		
/organism="Janthinobacterium sp. J3"		
/mol_type="genomic DNA"		
/strain="J3"		
/db_xref="taxon:213804"		
202. .609		
/country="Japan"		
202. .609		
/gene="ORFUI"		
202. .609		
/gene="ORFUI"		
/function="unknown"		
gene		
CDS		
/codon_start=1		
/transl_table=11		
/product="hypothetical protein"		
4765. .5106		
/gene="ORFU6"		
4765. .5106		
/gene="ORFU6"		
/function="unknown"		
/codon_start=1		
/transl_table=11		
/product="hypothetical protein"		
4765. .5106		
/gene="ORFU6"		
4765. .5106		
/gene="ORFU6"		
/function="unknown"		
/codon_start=1		
/transl_table=11		
/product="hypothetical protein"		
4765. .5106		
/gene="ORFU6"		
4765. .5106		
/gene="ORFU6"		
/function="unknown"		
/codon_start=1		
/transl_table=11		
/product="hypothetical protein"		
4765. .5106		
/gene="ORFU6"		
4765. .5106		
/gene="ORFU6"		
/function="unknown"		
/codon_start=1		
/transl_table=11		
/product="hypothetical protein"		
4765. .5106		
/gene="ORFU6"		
4765. .5106		
/gene="ORFU6"		
/function="unknown"		
/codon_start=1		
/transl_table=11		
/product="hypothetical protein"		
4765. .5106		
/gene="ORFU6"		
4765. .5106		
/gene="ORFU6"		
/function="unknown"		
/codon_start=1		
/transl_table=11		
/product="hypothetical protein"		
4765. .5106		
/gene="ORFU6"		
4765. .5106		
/gene="ORFU6"		
/function="unknown"		
/codon_start=1		
/transl_table=11		
/product="hypothetical protein"		
4765. .5106		
/gene="ORFU6"		
4765. .5106		
/gene="ORFU6"		
/function="unknown"		
/codon_start=1		
/transl_table=11		
/product="hypothetical protein"		
4765. .5106		
/gene="ORFU6"		
4765. .5106		
/gene="ORFU6"		
/function="unknown"		
/codon_start=1		
/transl_table=11		
/product="hypothetical protein"		
4765. .5106		
/gene="ORFU6"		
4765. .5106		
/gene="ORFU6"		
/function="unknown"		
/codon_start=1		
/transl_table=11		
/product="hypothetical protein"		
4765. .5106		
/gene="ORFU6"		
4765. .5106		
/gene="ORFU6"		
/function="unknown"		
/codon_start=1		
/transl_table=11		
/product="hypothetical protein"		
4765. .5106		
/gene="ORFU6"		
4765. .5106		
/gene="ORFU6"		
/function="unknown"		
/codon_start=1		
/transl_table=11		
/product="hypothetical protein"		
4765. .5106		
/gene="ORFU6"		
4765. .5106		
/gene="ORFU6"		
/function="unknown"		
/codon_start=1		
/transl_table=11		
/product="hypothetical protein"		
4765. .5106		
/gene="ORFU6"		
4765. .5106		
/gene="ORFU6"		
/function="unknown"		
/codon_start=1		
/transl_table=11		
/product="hypothetical protein"		
4765. .5106		
/gene="ORFU6"		
4765. .5106		
/gene="ORFU6"		
/function="unknown"		
/codon_start=1		
/transl_table=11		
/product="hypothetical protein"		
4765. .5106		
/gene="ORFU6"		
4765. .5106		
/gene="ORFU6"		
/function="unknown"		
/codon_start=1		
/transl_table=11		
/product="hypothetical protein"		
4765. .5106		
/gene="ORFU6"		
4765. .5106		
/gene="ORFU6"		
/function="unknown"		
/codon_start=1		
/transl_table=11		
/product="hypothetical protein"		
4765. .5106		
/gene="ORFU6"		
4765. .5106		
/gene="ORFU6"		
/function="unknown"		
/codon_start=1		
/transl_table=11		
/product="hypothetical protein"		
4765. .5106		
/gene="ORFU6"		
4765. .5106		
/gene="ORFU6"		
/function="unknown"		
/codon_start=1		
/transl_table=11		
/product="hypothetical protein"		
4765. .5106		
/gene="ORFU6"		
4765. .5106		
/gene="ORFU6"		
/function="unknown"		
/codon_start=1		
/transl_table=11		
/product="hypothetical protein"		
4765. .5106		
/gene="ORFU6"		
4765. .5106		
/gene="ORFU6"		
/function="unknown"		
/codon_start=1		
/transl_table=11		
/product="hypothetical protein"		
4765. .5106		
/gene="ORFU6"		
4765. .5106		
/gene="ORFU6"		
/function="unknown"		
/codon_start=1		
/transl_table=11		
/product="hypothetical protein"		
4765. .5106		
/gene="ORFU6"		
4765. .5106		
/gene="ORFU6"		
/function="unknown"		
/codon_start=1		
/transl_table=11		
/product="hypothetical protein"		
4765. .5106		
/gene="ORFU6"		
4765. .5106		
/gene="ORFU6"		
/function="unknown"		
/codon_start=1		
/transl_table=11		
/product="hypothetical protein"		
4765. .5106		
/gene="ORFU6"		
4765. .5106		
/gene="ORFU6"		
/function="unknown"		
/codon_start=1		
/transl_table=11		
/product="hypothetical protein"		
4765. .5106		
/gene="ORFU6"		
4765. .5106		
/gene="ORFU6"		
/function="unknown"		
/codon_start=1		
/transl_table=11		
/product="hypothetical protein"		
4765. .5106		
/gene="ORFU6"		
4765. .5106		
/gene="ORFU6"		
/function="unknown"		
/codon_start=1		
/transl_table=11		
/product="hypothetical protein"		
4765. .5106		
/gene="ORFU6"		
4765. .5106		
/gene="ORFU6"		
/function="unknown"		
/codon_start=1		
/transl_table=11		
/product="hypothetical protein"		
4765. .5106		
/gene="ORFU6"		
4765. .5106		
/gene="ORFU6"		
/function="unknown"		
/codon_start=1		
/transl_table=11		
/product="hypothetical protein"		
4765. .5106		
/gene="ORFU6"		
4765. .5106		
/gene="ORFU6"		
/function="unknown"		
/codon_start=1		
/transl_table=11		
/product="hypothetical protein"		
4765. .5106		
/gene="ORFU6"		
4765. .5106		
/gene="ORFU6"		
/function="unknown"		
/codon_start=1		
/transl_table=11		
/product="hypothetical protein"		
4765. .5106		
/gene="ORFU6"		
4765. .5106		
/gene="ORFU6"		
/function="unknown"		
/codon_start=1		
/transl_table=11		
/product="hypothetical protein"		
4765. .5106		
/gene="ORFU6"		
4765. .5106		
/gene="ORFU6"		
/function="unknown"		
/codon_start=1		
/transl_table=11		
/product="hypothetical protein"		
4765. .5106		
/gene="ORFU6"		
4765. .5106		
/gene="ORFU6"		
/function="unknown"		
/codon_start=1		
/transl_table=11		
/product="hypothetical protein"		
4765. .5106		
/gene="ORFU6"		
4765. .5106		
/gene="ORFU6"		
/function="unknown"		
/codon_start=1		
/transl_table=11		
/product="hypothetical protein"		
4765. .5106		
/gene="ORFU6"		
4765. .5106		
/gene="ORFU6"		
/function="unknown"		
/codon_start=1		
/transl_table=11		
/product="hypothetical protein"		
4765. .5106		
/gene="ORFU6"		
4765. .5106		
/gene="ORFU6"		
/function="unknown"		
/codon_start=1		
/transl_table=11		
/product="hypothetical protein"		
4765. .5106		
/gene="ORFU6"		
4765. .5106		
/gene="ORFU6"		
/function="unknown"		
/codon_start=1		
/transl_table=11		
/product="hypothetical protein"		
4765. .5106		
/gene="ORFU6"		
4765. .5106		
/gene="ORFU6"		
/function="unknown"		
/codon_start=1		
/transl_table=11		
/product="hypothetical protein"		
4765. .5106		
/gene="ORFU6"		
4765. .5106		
/gene="ORFU6"		
/function="unknown"		
/codon_start=1		
/transl_table=11		
/product="hypothetical protein"		
4765. .5106		
/gene="ORFU6"		
4765. .5106		
/gene="ORFU6"		
/function="unknown"		
/codon_start=1		
/transl_table=11		
/product="hypothetical protein"		
4765. .5106		
/gene="ORFU6"		
4765. .5106		
/gene="ORFU6"		
/function="unknown"		
/codon_start=1		
/transl_table=11		
/product="hypothetical protein"		
4765. .5106		
/gene="ORFU6"		
4765. .5106		
/gene="ORFU6"		
/function="unknown"		
/codon_start=1		
/transl_table=11		
/product="hypothetical protein"		
4765. .5106		
/gene="ORFU6"		
4765. .5106		
/gene="ORFU6"		
/function="unknown"		
/codon_start=1		
/transl_table=11		
/product="hypothetical protein"		
4765. .5106		
/gene="ORFU6"		
4765. .5106		
/gene="ORFU6"		
/function="unknown"		
/codon_start=1		
/transl_table=11		
/product="hypothetical protein"		
4765. .5106		
/gene="ORFU6"		
4765. .5106		
/gene="ORFU6"		
/function="unknown"		
/codon_start=1		
/transl_table=11		
/product="hypothetical protein"		
4765. .5106		
/gene="ORFU6"		
4765. .5106		
/gene="ORFU6"		
/function="unknown"		
/codon_start=1		
/transl_table=11		
/product="hypothetical protein"		
4765. .5106		
/gene="ORFU6"		
4765. .5106		
/gene="ORFU6"		
/function="unknown"		
/codon_start=1		
/transl_table=11		
/product="hypothetical protein"		
4765. .5106		
/gene="ORFU6"		
4765. .5106		
/gene="ORFU6"		
/function="unknown"		
/codon_start=1		
/transl_table=11		
/product="hypothetical protein"		
4765. .5106		
/gene="ORFU6"		
4765. .5106		
/gene="ORFU6"		
/function="unknown"		
/codon_start=1		
/transl_table=11		
/product="hypothetical protein"		
4765. .5106		
/gene="ORFU6"		
4765. .5106		
/gene="ORFU6"		
/function="unknown"		
/codon_start=1		
/transl_table=11		
/product="hypothetical protein"		
4765. .5106		
/gene="ORFU6"		
4765. .5106		
/gene="ORFU6"		
/function="unknown"		
/codon_start=1		
/transl_table=11		
/product="hypothetical protein"		
4765. .5106		
/gene="ORFU6"		
4765. .5106		
/gene="ORFU6"		
/function="unknown"		
/codon_start=1		
/transl_table=11		
/product="hypothetical protein"		
4765. .5106		
/gene="ORFU6"		
4765. .5106		
/gene="ORFU6"		
/function="unknown"		
/codon_start=1		
/transl_table=11		
/product="hypothetical protein"		
4765. .5106		
/gene="ORFU6"		
4765. .5106		
/gene="ORFU6"		
/function="unknown"		
/codon_start=1		
/transl_table=11		
/product="hypothetical protein"		
4765. .5106		
/gene="ORFU6"		
4765. .5106		
/gene="ORFU6"		
/function="unknown"		
/codon_start=1		
/transl_table=11		
/product="hypothetical protein"		
4765. .5106		
/gene="ORFU6"		
4765. .5106		
/gene="ORFU6"		
/function="unknown"		
/codon_start=1		
/transl_table=11		
/product="hypothetical protein"		
4765. .5106		
/gene="ORFU6"		
4765. .5106		
/gene="ORFU6"		
/function="unknown"		
/codon_start=1		
/transl_table=11		
/product="hypothetical protein"		
4765. .5106		
/gene="ORFU6"		
4765. .5106		
/gene="ORFU6"		
/function="unknown"		
/codon_start=1		
/transl_table=11		
/product="hypothetical protein"		
4765. .5106		
/gene="ORFU6"		
4765. .5106		
/gene="ORFU6"		
/function="unknown"		
/codon_start=1		
/transl_table=11		
/product="hypothetical protein"		
4765. .5106		
/gene="ORFU6"		
4765. .5106		
/gene="ORFU6"		
/function="unknown"		
/codon_start=1		
/transl_table=11		
/product="hypothetical protein"		
4765. .5106		
/gene="ORFU6"		
4765. .5106		
/gene="ORFU6"		
/function="unknown"		
/codon_start=1		
/transl_table=11		
/product="hypothetical protein"		
4765. .5106		
/gene="ORFU6"		
4765. .5106		
/gene="ORFU6"		
/function="unknown"		
/codon_start=1		
/transl_table=11		
/product="hypothetical protein"		
4765. .5106		
/gene="ORFU6"		
4765. .5106		
/gene="ORFU6"		
/function="unknown"		
/codon_start=1		
/transl_table=11		
/product="hypothetical protein"		
4765. .5106		
/gene="ORFU6"		
4765. .5106		
/gene="ORFU6"		
/function="unknown"		
/codon_start=1		
/transl_table=11		
/product="hypothetical protein"		
4765. .5106		
/gene="ORFU6"		
4765. .5106		
/gene="ORFU6"		
/function="unknown"		
/codon_start=1		
/transl_table=11		
/product="hypothetical protein"		
4765. .5106		
/gene="ORFU6"		
4765. .5106		
/gene="ORFU6"		
/function="unknown"		
/codon_start=1		
/transl_table=11		
/product="hypothetical protein"		
4765. .5106		
/gene="ORFU6"		
4765. .5106		
/gene="ORFU6"		
/function="unknown"		
/codon_start=1		
/transl_table=11		
/product="hypothetical protein"		
4765. .5106		
/gene="ORFU6"		
4765. .5106		
/gene="ORFU6"		
/function="unknown"		
/codon		

/protein_id="BAC56734.1"
/db_xref="GI:28201199"
/translation="MYGPAWYKVDNMRNETHIALILVATLAGNBSPGLECTWK
ADGLVPMWTFPRGEATEAMGVIEKVDTTNGNVKVTYKDLMKGSALAFITVNRNTA
NNMYTLHRIR"
5647..6171
/gene="ORFU7"
5647..6171
/gene="ORFU7"
/function="putative cold shock protein"
/codon_start=1
/transl_table=11
/product="cold shock protein"
/protein_id="BAC56735.1"
/db_xref="GI:28201200"
/translation="MELQGRVIEWDERGFGFIQNGTSETFLHISALRKTGRPLL
GSLVTEIRDDGRGLQVAEALIGTKSLRQSSGFSFSAVLGLVLLMFGVGVYRF
SNGQISASVYKIVLARDALRPHPEFQCTPEKSCSKMASCARALFHQRCNVKMD
GDGDIPEBQWCN"
6306..6818
/gene="ORFU8"
6306..6818
/gene="ORFU8"
/function="unknown"
/codon_start=1
/transl_table=11
/product="hypothetical protein"
/protein_id="BAC56736.1"
/db_xref="GI:28201201"
/translation="MKIFASLALILLATVPSTAQVATPASEARFTLLRLDGGWRM
SGDVRGQVYTSMAAALQGTFTENRMKDQVPAQVSVFIFGPKASQTVIAHMD
NHGAKYSIPHGTCQITGNSQLLFPVQSGNFRNTMYHGDTSSTVLESADPDGSK
HFAEYEVKN"
6950..7378
/gene="ORFU9"
6950..7378
/gene="ORFU9"
/function="putative transposase ORFA protein"
/codon_start=1
/transl_table=11
/product="transposase"
/protein_id="BAC56737.1"
/db_xref="GI:28201202"
/translation="MARALLPEDLWLSLIAHLFAHRSPKGRPRIDDRAPTGLFV
LKTGIPWEYIPRELGGSGMTWRLHEWMLAGWQRIHETILRLRHEHQIMWDRAS
VDAASVAPAGGHTGRNPTDRGLGCKHLWLSVGFHW"
7348..7629
/gene="ORFU10"
7348..7629
/gene="ORFU10"
/function="putative transposase"
/codon_start=1
/transl_table=11
/product="transposase"
/protein_id="BAC56738.1"
/db_xref="GI:28201203"
/translation="MVDERGFPLVAQISGAQVHDSRFLPLVESVPAVKLAGARXR
FGKUHADRAYSRAHRAWLRRIARIGYVESRERLGRWRWVERT"
complement(7862..8617)
/gene="carR"
complement(7862..8617)
/gene="carR"
/function="putative transcriptional regulator of car
operon"
/codon_start=1
/transl_table=11
/product="regulatory protein"
/protein_id="BAC56739.1"
/db_xref="GI:28201204"
/translation="MDNITSYLDQAPKASARAVKETSHEHVERQSRQVVEIAKA
IANGLLRFPAGQRLVEADLTKELGISRNPLREAFRLVAREGLLQVEPRGAIVARL
SRKELVDYFVRECLSEAAFLAENDKSPFRNYVLEASLNPSTTLPNVSDYLE
DNRFHSLILKAGNSHLPKLLDQLQFPITQSAPFRMYDANIYQNSTAEHYEILQAIL

AGDAARAAVRAAHRHTSKLYSLPDHFFNSD"
complement(8620..8880)
/gene="ORFU12"
complement(8620..8880)
/gene="ORFU12"
/function="unknown"
/codon_start=1
/transl_table=11
/product="hypothetical protein"
/protein_id="BAC56740.1"
/db_xref="GI:28201205"
/translation="MILQVEFLDNLNAILFPCDILGLSFEQAASVCYFRKIALQNC
SNYSFYNFRYQNFILKLDKWNTRMTVSNHGLVKNKLGK"
9141..10130

gene	Query Match	88.0%;	Score	871.6;	DB 1;	Length	20017;			
CDS	Best Local Similarity	92.5%;	Pred. No.	1.6e-224;						
	Matches	916;	Conservative	0;	Mismatches	74;	Indels	0;	Gaps	0;

QY 1 ATGTACCACTCAAAATGAAGGCAAGCGCCAGGACCTCGCGCTCAGGGAAGAGCGCTG 60
DB 14091 ATGTACCACTCAAAATGAAGGCAAGCGCCAGGACCTCGCGCTCAGATAAGATTGA 14150
QY 61 TTGGTCTCAGCACCTTGCTTAATGGTATCGGATTTCCGTACAGTGTGCATCGGAGGTTGC 120
DB 14151 TTGGTCTCGCGCTCGCAACGGTATCGGACTTCGTACAGTGTGCTCGGAGGTTGC 14210
QY 121 GGAGTATCAAAATTCGAGTTACTCGAAGGGAATGTCCTCAATCAATGTGCCCGCATGCTCCA 180
DB 14211 GGAGTATCAAAATTCGAGTTACTTCGAAGGACCGTCCCAATCAATGTGCCCGCATGCTCCA 14270
QY 181 GGACTTTCTTCGCGAGATCGTGAGAAGGGCAACCGCATCTTGTGATGCCAGTGGGTTGCG 240
DB 14271 GGACTTTCTTCGAGATCGTGAGAAGGGCAACCGCTCACCTTGTGATGCCAGTGTATTGG 14330
QY 241 TTCTCAGACCTGCGGATCAAAAGTCGAGTCGAGCAAGTAGTCCGACCAAGCAATGCCAATC 300
DB 14331 CTTTCAGACCTGCGCATCAAAAGTTGAGTCCGAGTAAAGTATATATACCAAGCATTC 14390
QY 301 TCAAGATGGAAGCGGAAGTTGTTGAGTCCGCGCTAACTCATGACTCTGCTCCGCTG 360
DB 14391 TCGAATATGGAGCGGAAGTTGTTGCGTCCGCGGCTAACTCATGACTCTGCTATCTGTA 14450
QY 361 CGATTACCACTGATGGCCAGCAAAATTCCTCCCGGCCAGTTGCTGCTAGTAGAGCA 420
DB 14451 AAATTACGACTGATGTCAGCAAAATTCCTCCCGGCCAGTTGCTGCTAAATTGAGCA 14510
QY 421 GAGCAGTTCCGAGCGCTGTTGCGCATATTCATGCGCAATTTAAAGAACCCCGAAGGC 480
DB 14511 GAGCAGTTCCGAGCGCTGTTGCGCATATTCATGCGCAATTTAAAGAACCCCGAAGGC 14570
QY 481 ATATGGAGTTCTATATTAAGAGGATACCCAGAGCATTTAGTCTTGGCTTTTGA 540
DB 14571 TTCTGGAGTTCTATATTAAGAGGATACCCAGAGCATTTAGTCTTGGCTTTTGA 14630
QY 541 AATAGAAAGAGGCGCTCGTCTATTTTGGAGGACCAATGGGCACATCTTCTTCCGT 600
DB 14631 AATAGAAAGAGGCGCTCGTCTATTTTGGAGGACCAATGGGCACATCTTCTTCCGT 14690
QY 601 CCAGGACCGCGCCGAAAGAGTCTTTGCAATTTGCGCGCGTGGCGGCTCTCGTATCGGCC 660
DB 14691 CCAGGACCGCGCCGAAAGAGTCTATGCAATTTGCGCGCGTGGCGGCTCTCGTATCGGCC 14750
QY 661 GCTATTGACGCGCTCGATGCGCAACAGACCGGTAAAGTTCTTACCGCTCA 720
DB 14751 GCTATTGACGCGCTCGATGCGCAACAGACCGGTAAAGTTCTTACCGCTCA 14810
QY 721 AGAATCTCCGCGGACGCTGTTTCGGTGGATCGATATCGACATCGATGAGGACAAGCTTGA 780
DB 14811 AGAATCTCCGCGGACGCTGTTTCGGTGGATCGATATCGACATCGATGAGGACAAGCTTGA 14870
QY 781 GTCGTCAGGCAAGTACGGAAGACCGGATAGCCTTTGCGAAGGCCCGCCTGTTTATT 840

```

Db      14871  GTCTGCAGGCGAGTTACGGAAGACACGAGTAGCCTTTGGCAGGGGCCCATCGGTTTATC 14930
QY      841   CATCAGGTTGTCAGCGAGCGCTGCTTGAACCCCTACCGCAATACGAATTTATCTTGCC 900
Db      14931  CATCAAGTTGTGACGAGCGCTGCTCGAAACCCCTACCGGAATACGAATTTATCTTGCC 14990
QY      901   GGTCCACCGCTATGTCGACGCTACTGCTCGTATGCTGCTCGCAAGGCTGTTCACGC 960
Db      14991  GGTCCACCGCGATGTTGGACGCGCACTGCTCGTATGCTGCTCGCAAGGCTGTTCACGC 15050
QY      961   GATCAAAATTCATTTTGACGCAATTTTCTAA 990
Db      15051  GATCAAAATTCATTTTGACGCAATTTTCTAA 15080

RESULT 7
AF065891
LOCUS      AF065891               8620 bp    DNA        linear    BCT 14-NOV-1998
DEFINITION Ralstonia eutropha JMP134 phenol hydroxylase gene cluster, complete
sequence.
ACCESSION  AF065891
VERSION     AF065891.1
KEYWORDS   GI:3172505
SOURCE     Ralstonia eutropha
            Ralstonia eutropha
            Burkholderia; Betaproteobacteria; Burkholderiales;
            Burkholderiaceae; Ralstonia.
REFERENCE  1 (bases 1 to 8620)
            Ayoubi, P.J. and Harker, A.R.
            whole-cell kinetics of trichloroethylene degradation by phenol
            hydroxylase in a ralstonia eutropha JMP134 derivative
            Appl. Environ. Microbiol. 64 (11), 4353-4356 (1998)
JOURNAL   98013635
MEDLINE   9797289
FDBM      2 (bases 1 to 8620)
REFERENCE  Harker, A.R. and Ayoubi, P.J.
            Direct Submission
            Submitted (15-MAY-1998) Microbiology, Brigham Young University,
            P.O. Box 25108, Provo, UT 84602-5108, USA
FEATURES
            source
            1..8620
                /organism="Ralstonia eutropha"
                /mol_type="genomic DNA"
                /strain="JMP134"
                /db_xref="taxon:510"
                /chromosome="I"
            18..30
                /note="similar to TbuT, XylR, and DmpR binding sites"
            repeat_region
            49..61
                /rpt_type=inverted
            repeat_region
            56..68
                /note="similar to TbuT, XylR, and DmpR binding sites"
                /rpt_type=inverted
            56..68
                /bound_moiety="1HP"
            misc_binding
            191..218
                /evidence=not_experimental
            promoter
            284..1801
                /note="consensus sigma-54 promoter region"
            gene
            284..289
                /gene="phlK"
            RBS
            284..289
                /gene="phlK"
            CDS
            296..1801
                /gene="phlK"
                /codon_start=1
                /transl_table=11
                /product="putative hydroxylase component"
                /protein_id="AAC77380.1"
                /db_xref="GI:3172508"
                /translation="MALLERAANYDIARTNWTFTYVAESDLFFEVNAGQGVPMIEM
                GEYAMGAERMAFGAPGRMNATFGMLDENHGQLQYFPDYCPKDRQPDWAHK
                AYHTNQGAIAARTFDLPMRSRAIEIAVLTAFETGTNNQFLGLAADAAGDF
                TFASSLIQTDSESHAQIGGPALQILIANGRKEQAQQLVDVAIAARWLFSLLTGTG
                MDYATPLQHRKSEKFEKMTIEWIGQFERTLIDLGLPWYWDQMINFEDYQHAYQMG
                IWFVRPTVWNPAAQMTPCDRDLSEKYPGWNTDFGKAWDVIIDNLAGRRLTVPET
                LPIVRSQTPICAVPGNGNVRDYPLEYNGRTVHENSEIDRWYQODPVYRDHLTL
                VDRLAGHIQPPDLGGALRYNNLAPGEIGDDAHQYAVNVEAYRROEQNKAA"
                1817..2086
                /gene="phlL"
                1817..1821
                /gene="phlL"
                1826..2086
                /gene="phlL"
                /codon_start=1
                /transl_table=11
                /product="putative hydroxylase component"
                /protein_id="AAC77381.1"
                /db_xref="GI:3172507"
                /translation="MALPPTVISNFQYDFVLQLVAVDTENSMDEVAASAAAHSHSVGRRA
                POPGVVRVRSQGGQFYPRDARIGDTDIKPMESLEIFCDA"
                2085..2429
                /gene="phlM"
                2085..2089
                /gene="phlM"
                2094..2429
                /gene="phlM"
                /codon_start=1
                /transl_table=11
                /product="putative ferredoxin"
                /protein_id="AAC77382.1"
                /db_xref="GI:3172508"
                /translation="MNFQKVSITDELMEGDMAEYVDGHVIVLRPEGAPRAFQIC
                PHQDPLAEGKFGDGRVLMCRHQWTFDANTGKINGPGCRLAEYAVKVDGDDILIAVE
                GVEPLFANC"
                2438..2765
                /gene="phlN"
                2438..2441
                /gene="phlN"
                2449..2765
                /gene="phlN"
                /codon_start=1
                /transl_table=11
                /product="putative hydroxylase component"
                /protein_id="AAC77383.1"
                /db_xref="GI:3172509"
                /translation="MSKDHNTAEAYRNRRNVGVFLRASSITSGVIEAAQEDNPGKEIRV
                DDKLAYVRIDTDSLLILRRATLEDALGRPFMRSELVNLSFAGRIETDDYVRFYE
                KTL"
                2769..3777
                /gene="phlO"
                2769..2773
                /gene="phlO"
                2779..3777
                /gene="phlO"
                /codon_start=1
                /transl_table=11
                /product="putative hydroxylase component"
                /protein_id="AAC77384.1"
                /db_xref="GI:3172510"
                /translation="MTTQPDVLKPKLTKWHLAARRKPSYEIVSTNLHYTTDNPDP
                FELDPFEMAGWFKRHRAKPLKADNNAFRDPDELVYRTYNLQDQGETVYSGLLDQ
                PSEGRDMSLHSHWAGTLARLYTPARYLHALOMGSAVLTQMAPASTISNCAAYOTAD
                SLRWLTHTAYRTRELSOTFGDVGCTDERRRYEODPAWQMKLVKVEHALVAWDWAECE
                VAFSLVLRPAMEEAVLRGLGEAAHNGTILGLITDQALADYQRRHRHWAGALVEMALE
                TFGNRDLVLAGSIAVRLAPLADDAISAYCAALDPAPNARACAARVDFWDSIGLAGL"
                3818..4824
                /gene="phlP"
                3818..3821
                /gene="phlP"
                3826..4824
                /gene="phlP"
                /codon_start=1
                /transl_table=11
                /product="putative electron transfer component"
                /protein_id="AAC77385.1"
                /db_xref="GI:3172511"

```

```

/translation="WKYQISLEGGAFTVAABEDTLRGAALRAGNALPHECSVGGCGA
CPFDLVGLMSVWEAERGLSLACQRPUSDDCTIRVRCDSEYRPAVRAH
RAAEALLARALTDMSEFTFRVPQATEFRQYALLYPPRAPGARAYSMANLPNEB
IWKVIRVPVGGAGSNALFDQYIGDSVVLGDPYGHAYLRDSDARDIVCIAGGSLG
MLSVARGALGAGSRVHFYFARGQADLGDALAEKRAEDKRVTLVALSAPSTWK
GPTGFVHEVERNALTSLSYDFYFAGPPLMIEAQAALLMKHQVFPFQIIRDFRV"
5051..6837
/gene="phlR"
RBS
5051..5054
/gene="phlR"
CDS
5059..6837
/gene="phlR"
/codon_start=1
/transl_table=1
/product="putative transcriptional activator"
/protein_id="AAC77386.1"
/db_xref="GI:3172512"
/translation="WAKRHPGPKSGSPAQGHPRNQKRESPEAIRVPAIHDLAKRL
FAPQGRWLDQRMWMLHISLGALELIESLKGKTARGLITRIGQAGTRDAAMA
KRVKRALGTFDLGAPQLVSLGVHCEPVALDIDVQRHYFGDFYLVDSSEAEHI
AGGYIGNESVCNMLIGYACGYTSAPMGRIILWRETECGMHGVKRVVGVKPVSEDA
ODDLAPLRIGDFVWAPVDEAALPATSIARLSSAPENSFGVIGSAGPNTVCHMVN
KVAIDATVLFIGESGVKEVPANLHLISKADGPYAVNCASIPHELMSEELFQVE
RGFTGATSRAGRERADGGFLFDEVTGLSFTAGKLLRALQOGBIERVGDTRTK
VNRVVAATNVRRAVAGHFREDLFFRLNVFPIQVPLRERDDIPLMNWFLQRM
ARKDKHITGFRERADVFAYDFNENRELENMTIERAVILAEQDGLDLCHLSTGE
EYDTTAFMLKRSGSICRVSEASESPSAAREPGLAETEVAMLRAVAEANGLSR
AARVLGISRPTLAYRLQYGITPEAQ"
complement(6890..8309)
/gene="phlX"
complement(6890..8305)
/gene="phlX"
/codon_start=1
/transl_table=1
/product="putative transport facilitator"
/protein_id="AAC77387.1"
/db_xref="GI:3172513"
/translation="WPRASLSPENDKELVRGNYVRVAVALCAACGTPAGATDVFNL
EGFQVPSRANGTGAAPFUGPAAVMANPATIGMAENHGLHGLDMVTDIKAINAT
GSTAREGNNGNPPFAPEAPAVFYRHNDIALGAGIFABAGLGTQYGGSSFLTSNG
VQTGLDQFSLRLVLRIPPAVAYRPTDKLTGLASDAVWISLNLGLTLLVDTQIGTASQ
RVGSSLVLTLSVFLSGYLGFNSAPVGGGADAWGIGKLGITYDITDPTRVGLA
YNAKTRVGLRQGLAALVSTVAGNIPLSGEVVRDQFQWPAQFTVGIHQFNERLSA
ADYQVFWSSVMRDINVTFTRGANLDLSLPONVYRDTNVLGVAVYRDSAMTVRGGF
HYAQBAILPNGLLAVVPAIPPTTLSSGATYAFKDDAIDFALSVDLPTVGNASQPNL
AVPLIVRHSQNALVLAIRKFP"
complement(8306..8309)
/gene="phlX"
RBS
ORIGIN
Query Match
Best Local Similarity 13.2%; Score 130.8; DB 1; Length 8620;
Matches 448; Conservative 0; Mismatches 442; Indels 12; Gaps 4;
QY 53 AGAGCCTGTGTGCTCAGCACTTGCTAATGATGATCGGATTTCCGTACGAGTGTGCATCGG 112
DB 3884 ACACGCTGTGCGCGCGCCCTCGCGCGCGCATGCGCCCTGCGCACAGTSCAGCGTGG 3943
QY 113 GAGGTTCGCGAGTATCAATTCGATTCGAGGGAATGTCCAACTCAATGTGCGCGG 172
DB 3944 GCGGCTTCGCGCGGTGCCCTTCGACCTTGTGACGCGGTCTAAGAACTCGTCTGCGCTG 4003
QY 173 ATGCTCCAGACTTTCTTCGAGATCGTGAGAAGCGCAACGCCCATCTTGTGATCCCACT 232
DB 4004 AGCGCGCGCGCTGTCCGAGCGTGACCGCAAGCGTGGCAA---GTACCTTGGGTGCCAGT 4060
QY 233 CGGTTGCGCTCTCAGACTCGGATCAAGTGCAGTGCAGACAAAGTACGTCCCAAGA 292
DB 4061 CGCGCGCGGTAGCGACTGCAGATACGTGTGCGGTGCGACCAATCGTATCGCCCGCGG 4120
QY 293 TTCCAATCTCAAGAAATGGAAGGAGTTGTTGAGTCCGCGCGGTCACTCATGACCTGC 352
DB 4121 TCCGCGCGCATGCCGCGCAGCGGACTCTCTGCGACGCGCGCGGTGACCTCGACATGA 4180

```

```

353 TGTCCGTGCGATTACGCACTGATGGGCCAGCAAAATTTCTCCCGGCCAGTTCTTGCTAG 412
DB 4181 GCGAAATTCAGTTTCCCGGTTCCAGGCGCGACCGAGTTCCGGCCCGCCAGTACGCGTGC 4240
QY 413 TAGAGGCAGAGCAGTTCGCCAGGCGTGGTTCGCGCATATTCAATGGCGAATTTAAAGAAC 472
DB 4241 TCTACCGCGCCCGTGCACCGGGC---GCCCGCCCTATTTCGATGGCCACCTCCCAACG 4297
QY 473 CCGAAGCATATGGGATTTCTATTATAAGAGGTACCCACAGGACGATTTAGTCTTGGC 532
DB 4298 AGGAAGGCATCTGGAAGTTCGTGATCCGCGCGTGCCTGGCGGGGGCTGGCAGCAACGCGC 4357
QY 533 TTTTCGAATATAGAAAGAGGCGCTCGTCTATTATTTTCGCGGACCAATGGGACATCTT 592
DB 4358 TGTTCGACCAAGTTCGGAATCGGGACAGCGTCTGTCTCGACGGCGCGTACGCCACGCT 4417
QY 593 TCTTCGTTCAGGACCGGCCGAAAGAGTCTTTTCATTTGGCGCGTGGCGGCTCTCGT 652
DB 4418 ACCTTCGTAGGACAGTGCCTCGGACATTTCTGTCATCGCGGTGCTCGGGCTCGCGC 4477
QY 653 ATGCGGCGCTATTTCACGCGCGCTCGATGGCGGAAACAGACACAGCG---GTAAAGTTGT 709
DB 4478 CAATGCTGTTCGTTCGACGCGGAGCACTCGCCGTAGCGGTTTCGCGCGCGCTCCATCTT 4537
QY 710 TCTACGCGCTCAAGAACTCCCGCGCGAGCGTCTTCGCTGGATTCGATACGACA---TCGATG 766
DB 4538 TTTATGCGCGCCCGGGGCGAGGCTGACCTCGTGCCTCGACGCCCTTGGAAAAAATTGCGC 4597
QY 767 AGGACAGCTTGAGTTCGTCCAGGCACTTACGGAAGACACAGGATAGCTTTGCGACAGGC 826
DB 4598 AAGACAAAGCGGGTTACGCTGTCGTGCGCACTGTCGCGACCGGAGAGCACCTTGAAGGGC 4657
QY 827 CCATCGTCTTATTCATCAGTTTTCGACGCGCGTGTCTTGAACCCCTACCGGAATACG 886
DB 4658 CAACGGGTTTCGTTCAGGAGGAGTTCGAGGTAACCTGACAGCTTCCCTCGGCAGTACG 4717
QY 887 AATTTATCTTCGCGGTCCACCGCTATGTCGACGCTACTGTCGATGCTCTCTCGGCA 946
DB 4718 ACTTCTATTTTCGCGCGCGCCACTCATGATCGAAGCCATGTCAGGCGCTCTCTGATGACA 4777
QY 947 AG 948
DB 4778 AG 4779
RESULT 8
DB3068
LOCUS
DEFINITION Pseudomonas aeruginosa 5337 bp DNA linear BCT 06-FEB-1999
SUBMIT benzene monooxygenase ferredoxin, complete cds.
ACCESSION DB3068
VERSION DB3068.1 GI:1754622
KEYWORDS benzene monooxygenase oxygenase subunit; benzene monooxygenase ferredoxin; bmoA; bmoB; bmoC; bmoD1.
SOURCE Pseudomonas aeruginosa
ORGANISM Pseudomonas aeruginosa
REFERENCE 1 (sites)
AUTHORS Bacteria; Proteobacteria; Gammaproteobacteria; Pseudomonadales; Kitayama,A., Suzuki,E., Kawakami,Y. and Nagamune,T.
TITLE Gene organization and low regioispecificity in aromatic-ring hydroxylation of a benzene monooxygenase of Pseudomonas aeruginosa J1104
JOURNAL J. Ferment. Bioeng. 82, 421-425 (1996)
REFERENCE 2 (bases 1 to 5337)
AUTHORS Kitayama,A., Kawakami,Y. and Nagamune,T.
TITLE Gene organization and low regioispecificity in aromatic-ring hydroxylation of a benzene monooxygenase of Pseudomonas aeruginosa J1104
JOURNAL Unpublished
REFERENCE 3 (bases 1 to 5337)
AUTHORS Kitayama,A.

```


TITLE Direct Submission
JOURNAL Submitted (14-JAN-1996) Atsushi Kitayama, The University of Tokyo,
Department of Chemistry and Biotechnology; 7-3-1 Hongo, Bunkyo-ku,
Tokyo 113, Japan (E-mail: kitayama@tansei.cc.u-tokyo.ac.jp,
Tel: 81-3-3812-2111, Fax: 81-3-5684-8402)

FEATURES
Location/Qualifiers
1. 5337
/organism="Pseudomonas aeruginosa"
/mol_type="genomic DNA"
/strain="J1104"
/db_xref="taxon:287"
313. .317
324. .1826
/gene="bmoA"
324. .1826
/gene="bmoB"
/note="putative"
/transl_table=1
/product="benzene monooxygenase oxygenase subunit"
/protein_id="BAAL1761.1"
/db_xref="GI:1754623"
/translation="MAVLRNIDVDVARTNWTBKYVTEDELPPELGGSDIPMEKX
EAYDEPKYQTYEVKQREKDAVYVKAALERSKMFENADPGWQSVLKLHFGAIPS
GEYAASTAEARMRFPKAPGRNWTATLGSMDIEIRHAQLQYFFPHEYSKDRQFDWAHK
AFDNEWAAIASRHFDDIMWARDAISVIMLIRFQETFTNMQFLGLAADAEEAGDF
TFSSLISSIQDRESRHAQIGGPTLIQILIEGRKEEAQKQVDFIAFWARLPSVLTGPI
MDYITPLEHRSQSKFEMOEVI VEQFERSIHDLGDKWYWDIELEOLDQOHGHWLIG
VYVWEPVWNNPTAGVTPERDMLKPYKGNWDTWGHCMQVILDNLVGRTSLTLPET
LPIVCMNCLINTTPGNWNVQDYSLEYNRLHFGESEPDWIFPQEPERYAGHMTL
VDRFLGLIQMDLGGALYMDLAPSGSDDAHGYSVWEVYKQLRTKKAS"
1894. .1898
1906. .2172
/gene="bmoB"
1906. .2172
/gene="bmoB"
/codon_start=1
/transl_table=1
/protein_id="BAAL1762.1"
/db_xref="GI:1754624"
/translation="MALIPLISNFEQGFVLQVLPVDVTENTMDEVAAAAHHSVGRVRK
PRPGHIVRVKQSGSEELIPRTAKISEAGIKMECLFWEAPKH"
2214. .2217
2221. .2556
/gene="bmoC"
2221. .2556
/gene="bmoC"
/note="putative"
/codon_start=1
/transl_table=1
/product="benzene monooxygenase ferredoxin"
/protein_id="BAAL1763.1"
/db_xref="GI:1754625"
/translation="MAFKKVCSLDDLMEGEMESFEVDGQEVLLIWPPEGELKSFQIGC
PHQDIPLVGKFGDKVLICRAHLWQFDACSGKGINPGDCAQAQYVVKIEGDDVFDTE
GIQPLFSHS"
2612. .2616
2626. .3069
/gene="bmoD"
2626. .3069
/gene="bmoD"
/codon_start=1
/transl_table=1
/protein_id="BAAL1764.1"
/db_xref="GI:1754626"
/translation="MGLRVWLARLYSGSRGTAATAAVMMQGGVASVSKSTHEKKEETAM
TSNTEIGHKLNRRVGPVRAGELADAAEAAREDPKREIRVDDKRAYRIDTDDM
IIRRTTGALGRPDMPELEINLSSFAGRITQNDYVRFYFAKRV"

ORIGIN
Query Match 11.8%; Score 117.2; DB 1; Length 5337;
Best Local Similarity 52.8%; Pred. No. 1.8e-20;
Matches 325; Conservative 0; Mismatches 283; Indels 8; Gaps 3;

LOCUS E32813 5337 bp DNA linear PAT 31-JAN-2002
DEFINITION Process for producing m-hydroxyphenylacetic acid.
ACCESSION E32813
VERSION E32813.1 GI:18623943
KEYWORDS JP 2000069974-A/1.
SOURCE unidentified
ORGANISM unclassified
REFERENCE 1. (bases 1 to 5337)
AUTHORS Kitayama, A., Nagamune, T. and Shimizu, M.
TITLE Process for producing m-hydroxyphenylacetic acid
JOURNAL Patent: JP 2000069974-A 1 07-MAR-2000;
SUMITOMO CHEM CO LTD
COMMENT OS Pseudomonas aeruginosa J1104
PN JP 2000069974-A/1
PD 07-MAR-2000
PF 01-SEP-1998 JP 1998247139
PR
PI ATSUSHI KITAYAMA, TERUYUKI NAGAMUNE, MASATOSHI SHIMIZU PC
C12N15/09, C12P7/42//C12N1/21, (C12N15/09, C12R1:19), (C12P7/42, PC
C12R1:19),
PC (C12N1/21, C12R1:19), C12N15/00, (C12N15/00, C12R1:19) CC
FH Key Location/Qualifiers
FT CDS (324)..(1826)
FT CDS (3098)..(4123).

58 CTGTTGGTCTCAGCACTTGTCTAATCGTATCGGATTCCTCGATGTCATCGGAGGT 117
4283 CTCCTCGGTGCGCGCTCCAGGCCGCTGTCGGGTTCCTCGATGTCATCGGAGGT 4342
118 TCGGAGTATGCAAAATTCGAGTTACTCGAAGGGAATGTCCTAATCAATGTCGCGGATGCT 177
4343 TCGGAGTATGCAAAATTCGAGTTACTCGAAGGGAATGTCCTAATCAATGTCGCGGATGCT 177
178 CCAGGACTTTTCGCGAGATCGTGAGAGGCAACCGCCATCTTCGATGTCAGTGCCTGCT 237
4401 CCGGCTTACCGACCGTACCGAGCC--GCAACCGCTATCTAGCTGTGAGTGTG--CCGAGCC 4457
238 GCGCTCTCAGCTCGGATCAAGTCCGAGTGCAGGAGCAAGTACGTCCTCCACGATTCGA 297
4458 GCAAGACCGACTCCGATCAAGTGAAGCGCGCTGAGGAATATCGCCCCAAGGTACAG 4517
298 ATCTCAAGATGGAAGCGGAAGTTGTTGAGGTCGCGGCGCTAACTCATGACCTGCTGCTC 357
4518 CCGCGGTATTTGTCGCCCGGTTGTCGAGGCGTTACGGGTAAACGATGACATCTTGGAG 4577
358 GTGCGATTACGACATGATGGCGCAGCAATTTCTCCCGGCGAGTTCGCTTAGTAGAG 417
4578 TCCCGGTTCTCGGTGATGGCGCTCGCGAGTTCTCGCGGCGAGTACGCGCTGAT--G 4634
418 GCAGACAGTTGCCAGGCGTGTTCGCGCATATTCAATGCGCAATTTAAAGAACCCCGAA 477
4635 CAACTGCTGGCGTGAGGCTTCGCGGCTTACTCGATGTCACACGCGCAACCTCGG 4694
478 GGCATATGGAGTTCTATATTAAGAGGTACCCACAGGAGATTTAGTCTTGGCTTTTC 537
4695 GGAGAGTGCATTTCCAGATCCGCGGCTGTCCTCGTGGCGGAGGCTCGGGGCTCGGCTGTG 4754
538 GAAATAGAAAAAGAGCGCTGCTCTATTTTTGAACGGGACCAATGGGACATCTTTCTTC 597
4755 GGCATCTCAACACGAGGATTCGTCGAGTTGAGGCGCCCTATGGCTTGCATTG 4814
598 CGTCCAGGACCGGCGGAAAGAGTCTTTGATGTCGCGGCGGCTCGCGGCTCTCGATGCG 657
4815 AGAACCGAAGTCCGCGGACATCGTCTCGTGGCGGAGGCTCGGGGCTCGGCGCTGTG 4874
658 GCGCTATTGACGCG 673
4875 GTGTCGATCGCGCTG 4890

FEATURES		Location/Qualifiers	
source		1. .5337	
ORIGIN		/organism="unidentified"	
		/mol_type="genomic DNA"	
		/db_xref="taxon:32644"	
Query Match		11.88; Score 117.2; DB 6; Length 5337;	
Best Local Similarity		52.88; Pred. No. 1.8e-20;	
Matches 325; Conservative 0; Mismatches 283; Indels 6; Gaps 3;			
Qy	58	CTGTGCTCTCAGACACTTGCTATGATCGGATTTCCGTACGAGTGTGCATCGGAGGT	117
Db	4283	CTCTTGGGTCGGCGCTCCAGCGCGTGTCTGGGTCCCGTACGAGTGCATTCGGGGGC	4342
Qy	118	TGGGAGATGCAAAATCGAGTTACTCGAAGGAAATGTCCAATCAATGTGGCCGATGCT	177
Db	4343	TGGGAAATGTCGCTTCGAGTTGTGGAGGGGAAATGATGTTGTG--CCGAGGC	4400
Qy	178	CCAGGACTTTCTTCGCGAGATCGTGAAGAGGCAACCGCATCTTGCATGCCAGTCGCTT	237
Db	4401	CCCGGCTTACCGACCGTGACCGACGCC---GCAACGGCTATCTAGCCTGTCACTGTGTC	4457
Qy	238	CGGCTCTCAGACCTCGGATCAAGTCGAGTCGAGTCAAGTCAAGTACGATCCCAAGTCCA	297
Db	4458	GCCAAGACCGACCTCGGATCAAGTCAAGTCAAGTCAAGTCAAGTCAAGTCAAGTCAAG	4517
Qy	298	ATCTCAAGAAATGGAAGCGGAAGTGTGTAGGTCCGGCGCTAACTCATGACCTGCTGTCC	357
Db	4518	CCCCGCTATTTCGTGCCCCGTTTGTGAGGCGTTACGGTCAAGTCAAGTCAAGTCAAG	4577
Qy	358	GTGCGATTACGACTGATGCGGACGAAATTTCTCCCGGCGAGTTCGCTTCGCTTCGCT	417
Db	4578	TTCCGTTCTGCTGATGTCGCTTCTGCGGAGTCTGCTGCGGCGAGTACGCGCTGAT--G	4634
Qy	418	GCAGAGCAGTTGCCAGCGTGTTCGGGCATATTCAATGCGAATTTAAGACCCCGAA	477
Db	4635	CAACTGCTGCGTGGAGGTTGCGGCGTTACTGATGTCGACGCGGACCGACCTCGG	4694
Qy	478	GGCATATGAGGTTCTATTAAGAGGTTACCAAGAGGTTACCAAGAGGTTACCTGCTTTC	537
Db	4695	GGAGATGGGATTTCCAGATCCGCGGCTGCCAATGGCAAGAGTACCTGCGC	4754
Qy	538	GAAATAGAAAGAGCGCTCTCTATTTTTCAGCGGACCATGGGCACATCTTCTTC	597
Db	4755	GGCATCTCAACAGGAGGATTCGTGAGGTTGAGGCGCCCTATGCGCTTGCCTATTG	4814
Qy	598	CGTCCAGGACCGGCGGAAAGTCTTTGCAATTCGCGGCGGCTCGCGGCTCTCGTATCG	657
Db	4815	AGAACCGAAAGTCGCGGACATGCTCTGCTGCGGAGGCTCGGCGCTCGCGCTGTG	4874
Qy	658	GCGCTATTGACGCG 673	
Db	4875	GTGTCGATCGCGGTG 4890	
RESULT 10		4729 bp DNA linear BCT 26-SEP-2001	
AF282898		Burkholderia cepacia JS150 tbc2 gene cluster, complete sequence.	
DEFINITION		Burkholderia cepacia JS150 tbc2 gene cluster, complete sequence.	
ACCESSION		AF282898.1	
VERSION		GI:11878087	
KEYWORDS		Burkholderia cepacia	
SOURCE		Burkholderia cepacia	
ORGANISM		Burkholderia cepacia	
REFERENCE		Burkholderia cepacia	
AUTHORS		Khang,H.Y., Malinverni,J.C., Majko,M.M. and Kukor,J.J.	
TITLE		Genetic and functional analysis of the tbc operons for catabolism	
JOURNAL		of alkyl- and chloroaromatic compounds in Burkholderia sp. strain	
		JS150	
		Appl. Environ. Microbiol. 67 (10), 4805-4816 (2001)	

21455041

11571188

2 (bases 1 to 4729)

Khang,H.Y. and Kukor,J.J.

Direct Submission

Submitted (27-JUN-2000)

Biotech Center, Rutgers University, 59

Dudley Road, Cook College, New Brunswick, NJ 08901, USA

Location/Qualifiers

1. .4729

/organism="Burkholderia cepacia"

/mol_type="genomic DNA"

/strain="JS150"

/db_xref="taxon:292"

190..1695

/gene="tbc2A"

190..1695

/gene="tbc2A"

/function="catabolism of alkyl- and chloroaromatic compounds"

/codon_start=1

/transl_table=11

/product="tbc2A monoxygenase"

/protein_id="AAG40794.1"

/db_xref="GI:11878088"

/translation="MALLEAAADYDIARTTNWTPSYVTESELPDITMGQVPMETW

ETVPEYKTSYPEVVSIOREKADGYSKALERSRMFEDADPGLWLSILKAHYGAIAL

GEVDAAGARMAFRGAPGRNMAFGMDENRHOLOLYFPHDYCAKQDQPAHAK

AVHTNEGATAABSTFDDLFWRSALIDIALMLTAPETGTNNQFLGLAADAAGDE

TFASLISIQDTSRHAQIGGPAQLIILASGRKEQAKLIDIALPARAWRLFSLTGIS

MDYATPLHHRSPFEFTBWI VQGFERTLIDGLDLPFYWDQMINFDFYOHAYQMG

LPIVNNASQIPICAVPGNWTVDYPLDYGRTYHFNSEIDRWVFOQDPLRYDRDLTL

VDRLAGIQPPLMGLALQYMLNAPGCGDDAHYPMWEAYRNQRYQKAA"

1720..1980

/gene="tbc2B"

1720..1980

/gene="tbc2B"

/function="catabolism of alkyl- and chloroaromatic compounds"

/codon_start=1

/transl_table=11

/product="tbc2B monoxygenase"

/protein_id="AAG40795.1"

/db_xref="GI:11878089"

/translation="MALFPVVISNFPQYDFVQLVAVDTENTIDEVAAAHHSVGRVA

PQPKIVRVRRQGGEGQFYPRNRLADTDIKMEALEFIFCDA"

1988..2323

/gene="tbc2C"

1988..2323

/gene="tbc2C"

/function="catabolism of alkyl- and chloroaromatic compounds"

/codon_start=1

/transl_table=11

/product="tbc2C monoxygenase"

/protein_id="AAG40796.1"

/db_xref="GI:11878090"

/translation="MNFQKVCSDLDELWEGDMTEVEVAGHVILVVRPEGGBPRAQGIC

PHQDIPAEKGFDRVLMCRHAHQWTFDANTGKINPGCRLAELAYLKEGDDILVAVE

GVEPLFANC"

2343..2657

/gene="tbc2D"

2343..2657

/gene="tbc2D"

/function="catabolism of alkyl- and chloroaromatic compounds"

/codon_start=1

/transl_table=11

/product="tbc2D monoxygenase"

/protein_id="AAG40797.1"

/db_xref="GI:11878091"

/translation="MSIDNTAEAYRNRRVGLASLIITAGVIEAAQEDNPGKSIKMD

DKFAYVIDDGEILRLDEEALGPFKMSLEVLNLSFAGRIETIDYVRPYEK

gene TL" 2670. .3728
/gene="tbc2E"
CDS 2670. .3728
/gene="tbc2E"
/function="catabolism of alkyl- and chloroaromatic compounds"
/codon_start=1
/transl_table=11
/product="Tbc2E monooxygenase"
/protein_id="AAG40798.1"
/db_xref="GI:11878092"
/translation="MTQAEVLKPLKMTSLHARRKPESEYEIVSNLHYTNDPAP
FEIDPNEMAQKFNKNASPLHPDNAPRDELYRYTNMLQOQETVYFGLDQ
FSRGHDAUERWAGTLAKLYTPRYLFTLQMSAYLTLQAPASTINSNCAAYQAD
SLRWLTHAYTELSTDFSGDFGDEREWEQPAWQGRKLVHEHALVANDWAEFC
VALSVYRPAVEARLSEGEARHNGDITLLGLTDAQADARHWRWAGALVMALE
QPNREVITGWLAKWEPLADEAIVAYCSALPEAPAAQAAVREFTAPQPRVKDS
PGDIIQNPWPTTHNRR"
3725. .4729
/gene="tbc2F"
/gene="tbc2F"
/functions="catabolism of alkyl- and chloroaromatic compounds"
/codon_start=1
/transl_table=11
/product="Tbc2F monooxygenase"
/protein_id="AAG40799.1"
/db_xref="GI:11878093"
/translation="MKHQITIEGSAFSAVAEDDTLLRGALRAGIALPHECSVGGCGA
CRDLLGLVESIWEAPGLISERDRKRGKHLACQRPGLGDTIRVCDTIRVWSPG
RWAALCARRELTPDMSEFTFAVHAPAEFRPGVALLYLPPHAPGARAYSMNLPNADG
IWQVIRVPGGAGSNALFDVBIQOTVTLDDGYGHAHLRDDNARDIVICAGSGGLAP
MLSARGALAEQGNQRVHFYFGRSQPDLCMAALDDLVDNRLALSIVLSAPGPELG
WQGTGFVHAEVERVLVAPLDREFYFAGFPFMIEAVQALLMHTHRVPFEQMHDFRV"
ORIGIN
Query Match 11.0%; Score 109; DB 1; Length 4729;
Best Local Similarity 48.5%; Pred. No. 2.9e-18;
Matches 462; Conservative 0; Mismatches 470; Indels 21; Gaps 5;
Qy 53 AGACCTGTGGTCTCAGCACTTGCCTAATGATTCGGATTCGGTACGATGTCATCGG 112
Db 3783 ACACCTTGTCTCGAGGTGCGCTGCGCGCGGCATCGCCCTGCCCATGATGAGCGTGG 3842
Qy 113 GAGGTTCGGAGTATGCAATTCGAGTACTCGAAGGGAATGCCAATCAATGTGCGCGG 172
Db 3843 GCGCTGCGGCGCAATGCGCTTCGACTGCTCTCGGGCTCGTGGAAATCCATCTGCGCGG 3902
Qy 173 ATGCTCCAGGACTTCTTCGCGAGATCGTGAGAGGCAACCGCCATCTTTCATGCGCACT 232
Db 3903 AGGCGCTTGGCTCTCGAAGCGACGCGCAAGCGCGCAA---ACACCTCGCTGCCAAT 3959
Qy 233 GCGTTGCGCTCTCAGACTCGGATCAATTCGAGTTCGAGGACAGTACGTCCCAAGA 292
Db 3960 CGCGCCGCTGGCGATTGACAGTTCGCGTGCCTCGCAGACACTTACCGCCCGGTGG 4019
Qy 293 TTCCAATCTCAAGAAATGGAAGCGGAAGTTGTGAGGTCCGGGCGCTCACTCATGACCTGC 352
Db 4020 TGTCCCGCGCGCTGGCGCGCGTGTGCGAGGCGAGCGGAGCTACGCCGACATGA 4079
Qy 353 TGTCCGTGCGATTACGACTGATGGCGCAGCAATTCCTCCCGCGGCACTTTCGCTAG 412
Db 4080 CGCAATTACGCTTCGAGTACACGCGCGCGTGAATTCGCCCGGGCGCATGCGCGTGC 4139
Qy 413 TAGAGGCGAGACGATGTCGAGCGGTGTTCCGCGATATTCATGGCAATTTAAAGAAC 472
Db 4140 TGTATCCGCCCGCAGCGCGCTGGCGG---CGCGCTATTCCATGTCCAACTGCCCAAG 4196
Qy 473 CCGAAGGCATATGGAGTCTTATATTAAGAGGGTACCCACAGGACGATTAAGTCTTCGCG 532

Db 4197 CAGACGGCATCTGGCAGTTTGTGATTCGGCGCGGTTCGGCGGGGCGCGCAGCAACGGCG 4256
Qy 533 TTTTCGAAATAGAAAAGAGCGCGCTCGTCTATTATTTTACGCGGACCAATGGGCACATCTT 592
Db 4257 TGTTTGACCGAGTCCAGATCGGGCAGACCGTACCTCGACGGGCGCTATGGCCATGCC 4316
Qy 593 TCTTCGCTCCAGGACCGCGCGAAGAGTCTTTGATTTGGCGGCGGTGCGCGGCTCTCGT 652
Db 4317 ACCTGCGGACGACAAACGCGCGGACATCGTGTGATCGCTGGCGGCTCCGCGCTGCTC 4376
Qy 653 ATGCGCGCTATTTCAGCGCGCTCGAT---GCGGGAACAGACAGCGCGGTAAAGTTGT 709
Db 4377 CGATCTGTCTCGTCCCGCGGAGCGCTCGCGAGGAGGCGCGCAACGCTGCACTTCT 4436
Qy 710 TCTACCGCTCAAGAACTCCGCGCAGCGTGTTCGGTGGATCGATTCGACATCGATAGG 769
Db 4437 TCTACGGCGGGCGCTCGCAACCCGACCTCGCGCGCATCGCGCGCTCGATGACCTGTGTG 4496
Qy 770 ACAAGCTTGAGGTCTCGCAGCGAGTTA-----CGGAACAGCAGGATAGCCTTTGCG 820
Db 4497 GCGACAAACCGCTCGCGCTCTCGTGTGCTGTCGCGCGCGCGCGGAACTCGGCTGCG 4556
Qy 821 AAGGCGCCACTGTTTTTATTCATCAGGTGTGCGACGACGCTGCTTGAAACCCCTACCGG 880
Db 4557 AGGCGCGCGCGCTTCGTGTCATGCGAGGTGCAACGGGTGCTGTGTGCGCGCTGGACC 4616
Qy 881 AATAGAAATTTATCTTCGCGGTCCACCGCTTATGTCGACGCTACTCTCGGTATGCTGC 940
Db 4617 GCTTCAGTCTTATTTCCGCGCGCGCGCGCATGATCGAAGCGCTGCGGCAITTCGTGA 4676
Qy 941 TCGGCA---AGGCTGTTCCACGCGATCAATTCATTTTTCAGCGCATTTTCTAA 990
Db 4677 TGCACAGCATCGGTGCGGCTTTGAACAGATGCAATTCGACCGCTTCGCTCTAA 4729
RESULT 11
PPU04052
LOCUS
DEFINITION
Pseudomonas pickettii PK01 toluene-3-monooxygenase oxygenase
subunit (tbaA1) and (tbaA2) genes, (tbaU) gene, (tbaB) gene,
toluene-3-monooxygenase ferredoxin protein (tbaV) gene, and
toluene-3-monooxygenase oxidoreductase (tbaC) gene, complete cds.
U04052
VERSION
U04052.1 GI:1580790
KEYWORDS
SOURCE
ORGANISM
Ralstonia pickettii
Ralstonia pickettii
Bacteria; Proteobacteria; Betaproteobacteria; Burkholderiales;
Burkholderiaceae; Ralstonia.
REFERENCE
1 (bases 1 to 4993)
Byrne, A.M., Kukor, J.J. and Olsen, R.H.
Sequence analysis of the gene cluster encoding
toluene-3-monooxygenase from Pseudomonas pickettii PK01
Gene 154 (1), 65-70 (1995)
JOURNAL
MEDLINE
95172404
PUBMED
7867951
REFERENCE
2 (bases 1 to 4993)
Byrne, A.M. and Olsen, R.H.
Cascade regulation of the toluene-3-monooxygenase operon
(tbaA1-tbaA2) of Burkholderia pickettii PK01: Role of the tbaU
promoter (tbaU1) in the expression of its cognate activator, TbaU
J. Bacteriol 178 (21) (1996) in press
REFERENCE
3 (bases 1 to 4993)
Byrne, A.M.
Direct Submission
Submitted (06-DEC-1993) Armando M. Byrne, Department of
Microbiology & Immunology, University of Michigan Medical School,
5641 Medical Science Building II, 1301 West Catherine Street, Ann
Arbor, MI 48109-0620 USA
COMMENT
On Oct 2, 1996 this sequence version replaced gi:505045.
FEATURES
Location/Qualifiers
1..4993
/organism="Ralstonia pickettii"

/mol_type="genomic DNA"
 /strain="PK01"
 /db_xref="taxon:329"
 18. .62
 /note="putative TbuT-binding site"
 /citation=[2]
 /rpt_type=inverted
 193. .207
 215. .252
 /rpt_type=inverted
 243. .256
 280. .284
 293. .1798
 /gene="tbuA"
 293. .1798
 /gene="tbuA"
 /note="putative"
 /codon_start=1
 /transl_table=11
 /product="toluene-3-monooxygenase oxygenase subunit"
 /protein_id="AB09618.1"
 /db_xref="GI:505046"
 /translation="WALLIERAAWDIARTNWTSPSYVTESELPDIMGAGQVPMETW
 ETYDPSYTSYPEYISYIUREKADAGYSVKAALERSMPEDADPGMLILKAHYGAIAL
 GEYAAAEARFARPGMRNNAITPGMLDENRHRGQQLFYFPHDCAKDRQFQWAHK
 AYHTNEWGAIARSFFDLFWRSRAIDAIMLTFATFETGTMTMOFLGLADAAAGDF
 TPASLISDTSESHAQIIGGPALQILITASGRKEAOAKLVDIATARAWRLPSLITGNS
 MDYATLPHRKHSESEFWEWTIVGPFERTLDGLDLPKYWDQMINEDFYOHAYONG
 IHWERTIWNPAAGITDCDRDLBEKYPGNDITFGKAWDVIIDNLAKRELTVPET
 LPIVNMQLPICAVPGNGWIVKDIPLDYKGRTHFNSEIDRWVTFQDPLRYOHLTL
 VDRFLAGIQIPPLMGLQYMNLAPEGCEGDAAHYAWVEAYRNQRYQKKA"
 1815. .1818
 1823. .2083
 /gene="tbuU"
 1823. .2083
 /gene="tbuU"
 /codon_start=1
 /transl_table=11
 /product="unknown"
 /protein_id="AB09619.1"
 /db_xref="GI:505047"
 /translation="WALSPVISNFQYDFVLQIVAVDTENTIDEVAAAAHSHVGRVA
 PFGKIVVRROGGEQIFRNARLADTDIKPWEALFIFCDA"
 2082. .2085
 2091. .2426
 /gene="tbuB"
 2091. .2426
 /gene="tbuB"
 /codon_start=1
 /transl_table=11
 /product="unknown"
 /protein_id="AB09620.1"
 /db_xref="GI:505048"
 /translation="WNFOKVCSDLBELWEDMTEVEVAGHVIIVLVRPRGEPRAFQIG
 PHQDPLABGKFDGRVLMCRAHQWTFDANTGKINGPGCRLEALYALKVEGDDILVAVE
 GVEPLFANC"
 2435. .2439
 2446. .2760
 /gene="tbuV"
 2446. .2760
 /gene="tbuV"
 /note="putative"
 /codon_start=1
 /transl_table=11
 /product="toluene-3-monooxygenase ferredoxin protein"
 /protein_id="AB09621.1"
 /db_xref="GI:505049"
 /translation="WSIDNTAAYRNRRVGPVPLRASSITAGVIEAAQSDNPKSIRID
 DKLAYVRIPTDDELILRRATLEALGRFPKSELEVNLSFAGRIETDDIVRYPYEK
 TL"
 2763. .2767
 2773. .3762
 /gene="tbuA2"
 /mol_type="genomic DNA"
 /strain="tbaA2"
 /note="putative"
 /codon_start=1
 /transl_table=11
 /product="toluene-3-monooxygenase oxygenase subunit"
 /protein_id="AB09622.1"
 /db_xref="GI:505050"
 /translation="MTTQAEVLKPLKWSHLAARRRKSPSEYIEIVSTNLHYTTDNPDAP
 FELDNFMAOWKFNRRNASPLTHPDNAPRDFDELVYRYTNMLQDGGTYVYFGLLDQ
 FSRGHDAMLERKTWAGTLARLYTPVRYLFTLQMSAYLTQLPASTISNCAAYQTAD
 SLRWLTHTAYRTKELSQTFSDLGFGTDERRYWEQDPAQWQMKLVQKALVAWDAECF
 VALSLVRPAVEAEVLRSLGEAARHNGDTLLGLLTDQAOLADARHRWAGALVRVALE
 QSGNREVITGWLAKWEPLADEAIVAYCSALPEAPAAQARATAAVERFRHSGL"
 3817. .3820
 3827. .4831
 /gene="tbuC"
 3827. .4831
 /gene="tbuC"
 /note="putative"
 /codon_start=1
 /transl_table=11
 /product="toluene-3-monooxygenase oxidoreductase"
 /protein_id="AB09623.1"
 /db_xref="GI:505051"
 /translation="MKHQITIEGGSFSAVAEDTLLRGLRGGIALPHECSVGGCGA
 CFPDLISGVEISWPEAPGLSERDRKRGKHLACQSRPLGDCTIRVCDTDPVVPVSPG
 RWAVALAREELTPDMSEFTFAVHAPEAFPCQYALLYPHAPGARAYSMNLPNADG
 IQQVTRVPGGAGSNALFDQVEIGTQVTLDPGYGHARLDDNARDIVICLAGSGGLAP
 MUSVAGALAGGAQRVHFYFGRSQPDGLGAWAALDDLVGDNRLALSVVLSAPGPELG
 WQGPQTGFVHAEVERVLVAPLDRFEFFYFAGPPPMIEAVQALLMHTHRVPEQMHFDRFV
 "

QY	593	TCTTCGCTCCAGGACCGCGAAGAGTCTTTTCGATTTGGCGGCTGCGGGCTCTCGT	652
Db	4419	ACCTGCGCGACGAAACCGCGCGACATCGTGTGCATCGTGGCGGCTCCGACTGGCTC	4478
QY	653	ATGCGCGCGCTATTTCACGCGCTCGAT---GCGGAAACACAGAACCGGTAAGTTGT	709
Db	4479	CGATGCTCTCGTTCGCGCGGAGCGCTCGCGCAGGAGGCGCGAAACGCGTGCACTTCT	4538
QY	710	TCTACGGCTCAAGAACTCCGCGGAGCGTTCGGTGGATCGATACGACATCGATGAGG	769
Db	4539	TCTACGGCGGCGCTCGCAACCGGACCTCGCGCGCATCGCGCGCTGGATACCTGTTGG	4598
QY	770	ACAAAGCTTGAAGTCTCCAGGACAGTTA-----CGGAAGACACGATAGCCTTTGGC	820
Db	4599	GCGAACCGGCTCGCGCTCTCGTCTGCTGCTCGCGCGCGCGCGGAACTCGGCTGGC	4658
QY	821	AAGGCCCACTGGTTTATTATCATAGGTTTCGACGAGCGCTGTTGAACCCCTACGG	880
Db	4659	AGGCCCGGACCGGCTCTGATGATCGGAGGTGCAACCGGCTGCTGGTGGCGCGCTGAAC	4718
QY	881	AATACGAATTTATCTTCGCGGTCACCGCCTATGCTCGAGCTACTGTCGGTATGCTGC	940
Db	4719	GCTTCGAGTCTATTTCGCGCGCGCGCGCGCCCATGATCGAAGCCGTGCAGGCATGCTGA	4778
QY	941	TCGGCA---AGGTTTCCACGCGATCAAAATTCATTTGACGCATTTTCTAA	990
Db	4779	TCACACGATCGGTCGGTCTTGAACAGATGCATTTGACCGCTTCGCTCTAA	4831
RESULT 12			
LOCUS	E39081	5743 bp	DNA linear PAT 31-JAN-2002
DEFINITION	Gene encoding desulfurase.		
ACCESSION	E39081		
VERSION	E39081.1 GI:18628914		
KEYWORDS	JP 2000245477-A/1.		
SOURCE	Burkholderia sp.		
ORGANISM	Burkholderia sp.		
REFERENCE	Bacteria; Proteobacteria; Betaproteobacteria; Burkholderiales;		
AUTHORS	Burkholderiaceae; Burkholderia.		
TITLE	1 (bases 1 to 5743)		
JOURNAL	Kato, A. and Yanase, E.		
COMMENT	Gene encoding desulfurase		
	Patent: JP 2000245477-A 1 12-SEP-2000;		
	PETROLEUM ENERGY CENTER		
	OS Burkholderia sp.		
	PN JP 2000245477-A/1		
	PD 12-SEP-2000		
	PF 04-MAR-1999 JP 1999057446		
	PR		
	PI AKIO KATO, EIJI YANASE		
	PC C12N15/09, C10G32/00, C12N9/02, C12N9/90///(C12N15/09, C12R1:01),		
	PC C12N15/00,		
	PC (C12N15/00, C12R1:01)		
	CC		
	FH Key Location/Qualifiers		
	FT CDS (78)..(1103)		
	FT CDS (2499)..(3329)		
	FT CDS (3389)..(4744)		
	FT CDS (4777)..(5271)		
	FT CDS (5455)..(5742).		
FEATURES	Location/Qualifiers		
source	1..5743		
	/organism="Burkholderia sp."		
	/mol_type="genomic DNA"		
	/db_xref="taxon:36773"		
ORIGIN			
Query Match	10.9%; Score 107.8; DB 6;		
Best Local Similarity	49.8%; Pred. No. 6.2e-18;		
Matches 330; Conservative	0; Mismatches 327; Indels 6; Gaps 2;		

QY	40	TGCGGCTCAGGGAAGAGCCTGTTGGTCTCAGCACCTTGCTAAATGGTATCGGATTCGGTAC	99
Db	135	TGTTCCACTGATGAACAGCTACTCGCTGGCGGCTCAAGGCTGGCTTGGATTTCCCTAT	194
QY	100	GAGTGTGCATCGGAGGTTGCGAGATATGCAAAATTCGAGTTACTCGAAGGGAATGTCMAA	159
Db	195	GAAATGCAATCCGGATCGTGCAGCAGTTTCCCGCTTTTCACTGCTGAGGCGCATGTCAAG	254
QY	160	TCAATGTGCGCGATGCTCCAGGACTTTCTTCGCGAGATCGTGAGAGGGAACCGCCAT	219
Db	255	GATTTGTGTCGAATGCGCGGTTTGAATGCGAAGCTTCGTAATGCGGCA---TGCAC	311
QY	220	CTTGTGATGCGAGTGTGCGCTCTCAGACTTGCAGATCAAAAGTCCGAGTCGACAGGACAAG	279
Db	312	CTCGGCTGCAGACACACAGTAGTACTGTGCAATCAAGCTGCGCTTGAAGCGCGGC	371
QY	280	TAGTCTCCCAACGATTCCTCAATCTCAAGATGGAAGCGGAAGTTTGTGAGTTCGCGGCGCTA	339
Db	372	TATGTCCCAGAGTGATGCCGACGCTCAGCATGCGCGCTTGGTGCATGTGTCGCCACTG	431
QY	340	ACTCATGACCTGCTGCTCCGTGCGATTACGCACTGATGGCCAGCAAAATTCCTCCCGGC	399
Db	432	ACGCAAGACATGTTTCGAGTTCCAGTTTCGTACGGAGGCTCCAGTTCGATTTTTCCTGGC	491
QY	400	CAGTTCTGCTAGTAGAGGACAGCAGTTGCCAGGCGTGTTCGCGCATATTCATGGCG	459
Db	492	CAGTATGCGCTGCT---TCGCTTTCCTGCGGTGCGCACAGGCGCTACTCGATGAGC	548
QY	460	AATTTAAAGAACCCCGAAGGCATATGGGAGTTCTATATTAAGAGGCTACCCACAGGACGA	519
Db	549	AACCTGCCAATGGCATTTGCGAATTCGATTCATCGTCAAGCAACCGGCTGTTGT	608
QY	520	TTTAGTCTTGGTTTTTCGAAATAGAAAGAGCGCTCTATATTTTTCGCGGACCA	579
Db	609	GGAACCGCGTGTCTTTTCGATGTCCTGAAACGAGGTGATGCTATTGAGTTGGAGGTCCT	668
QY	580	ATGGGCACATCTTTCTTCGCTCCAGGACCGCGCGCAAGAGTCTTTTCGATTTGCGCGGT	639
Db	669	TACGGGACGCTTATCTGCGTACCGAACCAGGTAGAGGTGCTCTGCAATGCTGCGCGC	728
QY	640	GCGGCGCTCTGATGCGCGCGCTATTGCAAGCGCTTCGATGCGCGGAAACAGACAGCGG	699
Db	729	TCGGAGCTTTCTCCGATGTTGTCATCTCGCGCGCGCATGTGCAATCCGCGGCAATGACC	788
QY	700	GTA 702	
Db	789	GAA 791	
RESULT 13			
LOCUS	AB024945	24448 bp	DNA linear BCT 27-DEC-2000
DEFINITION	Alcaligenes faecalis phenanthrene degradative gene cluster (phnAb to phnE genes), complete cds.		
ACCESSION	AB024945		
VERSION	AB024945.1 GI:4586270		
KEYWORDS	3,4-dihydroxyphenanthrene dioxygenase; phnC; 2-carboxybenzaldehyde dioxygenase; phnI; 1-hydroxy-2-naphthoate dioxygenase; phnG; trans-2'-Carboxybenzalpyruvate hydratase-aldolase; glutathione-S-transferase; isomerase; dioxygenase small subunit; dioxygenase large subunit; dihydrodiol dehydrogenase; ferredoxin reductase; ferredoxin.		
SOURCE	Alcaligenes faecalis		
ORGANISM	Alcaligenes faecalis		
REFERENCE	Bacteria; Proteobacteria; Betaproteobacteria; Burkholderiales;		
AUTHORS	Alcaligenaceae; Alcaligenes.		
TITLE	1 (sites)		
JOURNAL	Kiyohara, H., Tabata, Y. and Takizawa, N.		
REFERENCE	A phenanthrene degradative gene cluster in Alcaligenes faecalis		
	2 (bases 1 to 24448)		
	Unpublished		
	AFK2		
	Takizawa, N., Tabata, Y., Fujimoto, T. and Kiyohara, H.		

TITLE Direct Submission
JOURNAL Submitted (14-MAR-1999) Noboru Takizawa, Okayama University of Science, Applied Chemistry; 1-1, Ridai-cho, Okayama 700-0005, Japan (E-mail: takn@bioeng.dac.ous.ac.jp, Tel: 81-86-256-9552, Fax: 81-86-256-9552)

FEATURES
Location/Qualifiers
1..24448
/organism="Alcaligenes faecalis"
/mol_type="genomic DNA"
/strain="AFK2"
/db_xref="taxon:511"
1718..2059
/gene="phnA"
1718..2059
/gene="phnB"
/codon_start=1
/transl_table=11
/product="ferredoxin"
/protein_id="BAA76320.1"
/db_xref="GI:4586271"
/translation="MTKSNATAGDITQVLPICAADSLAPQVACDVANLPLAIYNL DSEYLATSNICTHEVALLDGFEGELVCAHGGSFNRTGKSLPKERLQVFSV VKDGGQVFKI"
2083..3111
/gene="phnA"
2083..3111
/gene="phnA"
/codon_start=1
/transl_table=11
/product="ferredoxin reductase"
/protein_id="BAA76321.1"
/db_xref="GI:4586272"
/translation="MSFQSFQVRIGAGDTFVFCSTDETLAAAKAGLGFPPYCCQSG SCSSCFQALLEGVDKLNAGPLNAZARECGMLQCQSTPGSDCRKLRKPGVVPQ WITTRQHALVDVPLTQDMFQFQTEAPAEPLPGQVALLRLPGVVGTRVSMNPL NGGEVHFVVKRPGCGGTAVLFDLVRGDAIELEGPYGAALRITETGRDVVICGGS GLSPMLSLRGAVCNPAETERLLMFGITPLDHCADVAFGEFLKRLVFLFAIS DVNAESANDWGERGLIHEVLAKHIGPNQGYDFYFCGPPMTDAVHKLLVLRVPS QLHDFRFV"
3210..4043
/gene="phnB"
3210..4043
/gene="phnB"
/codon_start=1
/transl_table=11
/product="dihydrodicol dehydrogenase"
/protein_id="BAA76322.1"
/db_xref="GI:4586273"
/translation="MAWLEQSVFLTGGVAGLGRALVKRLVEGANVTVLDSNARGLD ELVESPKRVAGSGDVRLNADNRKAVELAVRPGKLDTFNRQHLGLCPQCRPAG RCQRSPDEIVGINLGYNGIKAAAPALVRSGSVILLSSAFYAGGGVLYTVAK HAAVLIKQAHELAPYRVNGVAPGIIASDLRGPXSLGMBQSIITSVPLADLVKDIA FIGRLSDTESYTGSYVYLASARNAPATGVIINCDDGGMGVRSLVGPASGGKLLKFG G"
4100..5413
/gene="phnA"
4100..5413
/gene="phnA"
/note="iron-sulfur protein"
/codon_start=1
/transl_table=11
/product="dioxxygenase large subunit"
/protein_id="BAA76323.1"
/db_xref="GI:4586274"
/translation="MSTNRTQDVESLVDVKSGLQAKRIFWDQIYDLIEIQLFGKQW QFLTHDSLIPNGDFTTGYMAEDEVIMRQDDSVKAFNLNCTRGRARLVAAEAAR AFSCYTHGWFGLDGLKVYPMEKELVSLDKSKFLGRLEIRYVESYRGYYGNFDEA PSLNDYLDGVKFLDGLIWMIDNGGI ELVPPSRSLNKNWTPAENFVGVAYHGWTHL ASLMALEGALAAIGNEAPNEGLGIOTSRFGHFGALMDVAPVAHVAKSAGTLRY EWPASRKPVKVEKLGAVGRI YGSHLNGTI FPNNSFLGLTNTFKLAVPRGPIQIEVT WAIIVEKMPALKNVAVGWNSTFGSAGMLEADDSDNMTWTONNREGMTREGKLSQ LGLDFDLDDFPFGVIGQSAIGETLTIEGFTGPRKRLWAQTGERS"
5487..6005

gene
CDS
gene
CDS
gene
CDS
gene

/gene="phnAd"
5487..6005
/gene="phnAd"
/note="iron-sulfur protein"
/codon_start=1
/transl_table=11
/product="dioxxygenase small subunit"
/protein_id="BAA76324.1"
/db_xref="GI:4586275"
/translation="MNEILPVRGSESEVVSTLYRRLAHLAESEYEMMOLHPDQY QAYFQQLRYKRYKLPKVAIDRFDLARIQFESGVQMSVDPPIRRLSLN IEVFKASSTGNQVFLNVFVRNRVDETTYSYGRREEVVRGEAGLQLLRVSLDE RELLGKVNPEFL"
6161..6751
/gene="phnD"
6161..6751
/gene="phnD"
/note="similar to 2-hydroxychromene-2-carboxylate isomerase"
/codon_start=1
/transl_table=11
/product="isomerase"
/protein_id="BAA76325.1"
/db_xref="GI:4586276"
/translation="MLSFYDFDPSFYSLASIELPFIQRYGISVSKYKPIDIACAKRA IGNVPSNRDMPVKLTHLSRDQLQWQRYGTPLKFPFSDRLNTGFFYAAAGEARA EYVRAPHLTWGMGOAPSGEPVLRSIASEGMNVDDFMQFTDSVDGANFYKQSIDEGI ARSVFGVPMVVICDEMWNWGNDRLDVDFVYIQRASN"
7714..8325
/gene="gst"
7714..8325
/gene="gst"
/codon_start=1
/transl_table=11
/product="glutathione-S-transferase"
/protein_id="BAA76326.1"
/db_xref="GI:4586277"
/translation="WIKKIYAGACAGSLAHIALHEAGLTFFNVVDLRTQRTDGRS LADNHSYVPLEDEQILTEVPVLYIADLPEKGLAPAGALPLIDCRSGSUS LTNFTSRFRCSCRSSEEVKASRAFLQLTWDAQLAGRTYLVEGQFTVADYILF VVAGTKFTGIDVQELTRLGSLYLGRIQAQPSVEAAMEAEGLK"
8403..9287
/gene="phnH"
8403..9287
/gene="phnH"
/codon_start=1
/transl_table=11
/product="trans-2'-carboxybenzalpyruvate hydratase-aldolase"
/protein_id="BAA76327.1"
/db_xref="GI:4586278"
/translation="WKIIPSDIRGVGIMETPATEDASNNQTOOSVNLSETKMVRFL NTAKIDILATITGEGGATLTWPEKLVDCVQTNKRSLPFGAVTILNTRDIERA RALVDIGVDGLFVGPRNMLAMDDKSIYRYQDARALPGVPMVAYDNPVAFKGIKSTE AYELAKIPEVVASKTCGPALAKDMRAVGEKRIILPLATAWLAAQELPELALAWS GSVCAPTNNMALSQAILAGDWQRAKELSEKCNMGVGVHVCGRFGGLYGLQHPDRSF AFCSSRPCKSGSSSATALP"
9406..10455
/gene="phnG"
9406..10455
/gene="phnG"
/codon_start=1
/transl_table=11
/product="1-hydroxy-2-naphthoate dioxxygenase"
/protein_id="BAA76328.1"
/db_xref="GI:4586279"
/translation="WTMSDHSKFTVEDVPPPELQSKYKDSIVFLNWSVPATSGKQI BEAQITWSTVLPITIQETAKQRATHVLRDLRLLMTNPKRLHVDEAGTGPLMDFQCV MPGRAGARHPMNAALFVQCTGGVKSIVDGKOCVMEPGDLILTPGCMWHEHNEKG DPITLVDVLMHLHFLTGTGFGPQARDVRPOLADSAFTAGIIVPKLDGSDHASRSY SPIFRYPWDVAVRALVASPAASDGRSVRYGNPLTGGPSMDLI DSTLIQLSGSKSTRK YKSSAAMCFVVEGHGVSIQIGETIENSHPDVF TMPANLWASHQAKGTARIFQVSTR EYIRLGLTFEAFGE"

QY	580	ATGGGCACATCTTTCTTCCTCAGGACCGCGCGAAGAGTCTTTGCAATGGCGCGGT	639
Db	2674	TACGGCGCGGCTTATCTGCGTACCGAAACCGGTAGAGATGCTGCTGCAATGGTGGCGG	2733
QY	640	GCCGGCTCTCGTATCGCGCGCTATTGACGCGCTCGATCGCGGAAACAGACAGCGG	699
Db	2734	TCCGGACTTCTCCGATGTGTCACTCTGCGCGGCGGAGTGTGCAATCGGCAATGACC	2793
QY	700	GTA 702	
Db	2794	GAA 2796	
RESULT 14			
LOCUS	AP157565	8154 bp	DNA linear BCT 11-MAR-2003
DEFINITION	Sphingopyxis macrogoltabida ring hydroxylating dioxygenase beta subunit (thnA2), ring hydroxylating dioxygenase alpha subunit (thnA1), 2,4-dihydroxydec-2-ene-1,10-dioic acid aldolase (thnF), 2-oxodec-4-ene-1,10-dioic acid hydratase (thnE), 4-(2-oxocyclohexyl)-2-hydroxy-but-2,4-dienoic acid hydrolase (thnD), cis-dihydrodiol dehydrogenase (thnB), 1,2-dihydroxyaphthalene dioxygenase ThnC (thnC), Rieske-type ferredoxin (thnA3), and ferredoxin reductase (thnA4) genes, complete cds.		
VERSION	AP157565 AF204963 AF498315		
KEYWORDS	AP157565.2 GI:23330200		
SOURCE	Sphingopyxis macrogoltabida		
ORGANISM	Sphingopyxis macrogoltabida		
REFERENCE	Bacteria; Proteobacteria; Alphaproteobacteria; Sphingomonadales; Sphingomonadaceae; Sphingopyxis.		
AUTHORS	1 (bases 1 to 8154)		
TITLE	Andujar, E., Hernaez, M.J., Kaschabek, S.R., Reineke, W. and Santero, E. Identification of an extradiol dioxygenase involved in tetralin biodegradation: gene sequence analysis and purification and characterization of the gene product		
JOURNAL	J. Bacteriol. 182 (3), 789-795 (2000)		
MEDLINE	20100771		
PUBMED	10633115		
REFERENCE	2 (bases 1 to 8154)		
AUTHORS	Hernaez, M.J., Andujar, E., Rios, J.L., Kaschabek, S.R., Reineke, W. and Santero, E.		
TITLE	Identification of a serine hydrolase which cleaves the alicyclic ring of tetralin		
JOURNAL	J. Bacteriol. 182 (19), 5448-5453 (2000)		
MEDLINE	2044192		
PUBMED	10986248		
REFERENCE	3 (bases 1 to 8154)		
AUTHORS	Hernaez, M.J., Floriano, B., Rios, J.J. and Santero, E.		
TITLE	Identification of a Hydratase and a Class II Aldolase Involved in Biodegradation of the Organic Solvent Tetralin		
JOURNAL	Appl. Environ. Microbiol. 68 (10), 4841-4846 (2002)		
PUBMED	12324329		
REFERENCE	4 (bases 1 to 8154)		
AUTHORS	Moreno-Ruiz, E., Hernaez, M.J., Martinez-Perez, O. and Santero, E.		
TITLE	Identification and Functional Characterization of Sphingomonas macrogoltabida Strain Ifa Genes Involved in the First Two Steps of the Tetralin Catabolic Pathway		
JOURNAL	J. Bacteriol. 185 (6), 2026-2030 (2003)		
MEDLINE	22505664		
PUBMED	12618469		
REFERENCE	5 (bases 1 to 8154)		
AUTHORS	Hernaez, M.J., Floriano, B. and Santero, E.		
TITLE	Sequence of thnE and thnD encoding for a hydratase and an aldolase involved in tetralin biodegradation		
JOURNAL	Unpublished		
REFERENCE	6 (bases 1 to 8154)		
AUTHORS	Andujar, E., Hernaez, M.J. and Santero, E.		
TITLE	Direct Submission		
JOURNAL	Submitted (09-JUN-1999) Genetica, Universidad de Sevilla, Ap. 1095, Sevilla 41080, Spain		
REFERENCE	7 (bases 1 to 8154)		

gene	10277..11773		
CDS	/gene="phn1"		
	10277..11773		
	/gene="phn1"		
	/codon_start=1		
	/transl_table=11		
	/product="2-carboxybenzaldehyde dehydrogenase"		
	/protein_id="BA076329.1"		
	/db_xref="GI:4586280"		
	/translation="MGSAAKSVKLLSSRLMSSRCRPTGCGAIRPRAAQAHRFPQPA		
	RFTVAFASQRLGSEPMHETVALPECNLFYGRFHVPGQGYGTLNPATQEDLG		
	PVSANAAADVAQAALAFQSWHETKPAVRAAALREFGNRLTHAEELGFLDAINC		
	GNPVREMAADVFAATQVDYFAGLVHEAKGDTLPWGDGVLNYSVREPGVVARIVAN		
	HLPLAAKLAAPAAGNTVLKAATAPLSAYRLATLLDDVPPGVNVLSGDTPCG		
	QALVDPLVKVLSLIGPATKSLRSAAEKVMSVLELGGKPLIICPADMDKALE		
	GAUQGMFTWAGQCGSRSFVHSYDRIKEMVOKISQILKRPDTPATMGL		
	VSKAQEKWMSFIDSAHKEGRLVCGVRPSDKALDNGFWFTATVADFVTPAMLFRE		
	YFGPVLAVIPWDEEESAAGWAGFRLDFNMDKSGSPRPSGIR"		
gene	13991..14899		
CDS	/gene="phnC"		
	13991..14899		
	/gene="phnC"		
	/codon_start=1		
	/transl_table=11		
	/product="3,4-dihydroxyphenanthrene dioxygenase"		
	/protein_id="BA076330.1"		
	/db_xref="GI:4586281"		
	/translation="WANYTELYLGISVSDAAANKVATEVMGMQLVDEGEADRFVLR		
	MDLWHRVIVHTGGDDDLAVYGVAVAGAELEQAAKLTAGIEVRGTRAEARERV		
	LGLKLQDPAGNAHEIFVSLVDGKHPFGKPTGDDQGVGHILIRSVDPERAT		
	MRFVAVGVTGSSDYLLKLTGDTVPRIPLKNNRHSVALMFGMPKRAHMLVYEQ		
	DMRLDGLTKVHKHGDIAVLQVSHNDGALSFYAANPSGLWMEPGSGVCAPAQOE		
Query Match	10.6%; Score 104.6; DB 1; Length 24448;		
Best Local Similarity	49.5%; Pred. No. 4.9e-17;		
Matches	328; Conservative 0; Mismatches 329; Indels 6; Gaps 2;		
QY	40	TCCGCTCAGGAAGAGCCTGTGCTTCAGACCTTGCTAATGTATCGGATTTCCGTAC	99
Db	2140	TGTTCCACTGATGAACGCTACTCTCGCGCGCTCAAGGCTGGCTTGCATTTCCCTAT	2199
QY	100	CAGTGTGATCGGAGGTTGCGAGTATGCAAAATTCGAGTTACTCGAAGGAATGTCCAA	159
Db	2200	GAATGTCAATCGGATCTGACGAGCTTTCGCTTTTCAGTTCTGGAGGGCGATGTCAAG	2259
QY	160	TCAATGTGCGCGGATGTCCTCAGGACTTCTTCGCGAGATCTGTAGAGAGGCAACCGGCAT	219
Db	2260	GATTTGTGTCGAATCGCGCGGTTTGAATCCGAAGCTCGTGAATCGGCA---TGCAC	2316
QY	220	CTTGATGCCAGTGTGCTCTCAGACCTGCGGATCAAAGTCGAGTCAGCAGCAAG	279
Db	2317	CTCGCTGCCAGACACACAGGTAGTACTGTCAATCAAAGCTCGCTTGAAGCCGGGC	2376
QY	280	TACGTCCCAACGATTCGAATCTCAAGATCGAAGCGGAAGTTGTTGAGGTCGCGCGCGCTA	339
Db	2377	TATGTCCCGCAGTGTGATTCGAGCGGTCAGCATCGCGGTTGTCGATGTCGCTGCTG	2436
QY	340	ACTCATGACCTGCTCGCTGCGATGATGCGACTGATGGCCAGCAATTTCTCTCCCGGC	399
Db	2437	AGCAGACATGTTCCAGTTCAGTTTCGTACGGAGGCTCCAGCTGAGTTTTCCTCGGC	2496
QY	400	CAGTCTCTGCTAGTAGAGCAGAGCAGTTGCCAGCGGTGTTGCGCGATATTCATATGGCG	459
Db	2497	CAGTATGCGCTGCT---TCGCTTGGCTGCGGTGTCGCGACAAGGCGCTACTCGATGAGC	2553
QY	460	AATTTAAAGACCCGAGAGGATATGGAGTTCTATATTAAGAGGTTACCCACAGGACGA	519
Db	2554	AACTCTGCCAATGGCAATGGCAATGGCAATTCATCGTCAAAACGCAACCCGGTGGTTGT	2613
QY	520	TTTAGTCTCTGCTTTTTCGAAATAGAAAGAGCGCTCGCTCTATTTTTCAGCGGACCA	579
Db	2614	GGAACCGGGTCTGTTTCGATGCTGCTGAAGCGAGTGAATGATGATTGAGGAGGTCTCT	2673

AUTHORS JOURNAL	Hernaez,M.J., Andujar,E. and Santero,E.	
	Direct Submission	
	Submitted (15-NOV-1999) Genetica, Universidad de Sevilla, Ap. 1095, Sevilla 41080, Spain	
	Sequence update by submitter	
REFERENCE AUTHORS JOURNAL	8 (bases 1 to 8154)	
	Hernaez,M.J., Floriano,B. and Santero,E.	
	Direct Submission	
	Submitted (01-APR-2002) Departamento de Ciencias Ambientales, Universidad Pablo de Olavide, ctra. Utrera Km 1, Sevilla 41013, Spain	
REMARK REFERENCE AUTHORS JOURNAL	Sequence update by submitter	
	9 (bases 1 to 8154)	
	Moreno-Ruiz,E., Hernaez,M.J., Martinez-Perez,O. and Santero,E.	
	Direct Submission	
REMARK COMMENT	Submitted (05-JUN-2002) Departamento de Ciencias Ambientales, Universidad Pablo de Olavide, ctra. Utrera Km 1, Sevilla 41013, Spain	
	Sequence update by submitter	
	On or before Sep 26, 2002 this sequence version replaced	
	gi:10567596, gi:5617851.	
FEATURES source	Location/Qualifiers	
	1..8154	
	/organism="Sphingopyxis macrogoltabida"	
	/mol_type="genomic DNA"	
gene	/strain="TPA"	
	/db_xref="taxon:33050"	
	complement (18..563)	
	/gene="thnA2"	
CDS	complement (18..563)	
	/gene="thnA2"	
	/codon_start=1	
	/transl_table=11	
gene	/product="ring hydroxylating dioxygenase beta subunit"	
	/protein_id="AA26444.1"	
	/db_xref="GI:23330205"	
	/translations="MTELAKERSASRMTADLIIIEIQQLFREARLLDAERYDEWVEM LAEDLHYFMFNETFRADKSNIEDLTAAYFNDRDDIIIRLARLKTGTAWSDPA TRYTHLIGNVELTEKGEYVYSAFLLCRGNERDEDIQVCKRDLWRRTAEGKL VQVTHLQNTLTKSLNAPL"	
CDS	complement (588..1982)	
	/gene="thnA1"	
	complement (588..1982)	
	/gene="thnA1"	
gene	/codon_start=1	
	/transl_table=11	
	/product="ring hydroxylating dioxygenase alpha subunit"	
	/protein_id="AA26443.1"	
CDS	/db_xref="GI:23330204"	
	/translations="METRMNVGTKEIHGGAHIEYDSLMLNDEGWSVRKVYDRDIYEE EMQIFARVWLFVAHESQIPNPGDFVTTYNGEDGVIVARQDKSVGFINSCEPHGKN VCHADGNARFVCYTHGWAFFDIGNLMGNAEEVYDPGDIKSEWQMPVAKVGK GLIFATPDGAPSLDYLGDYRWYLCILDSDEGSELVGGTWKFWI PMNWKVAZENP VGDAYTAAMTHDSAKAMTGMAPPIDFENTFHSSANGHGLEFLDGLDITALLGRP KVTEYFMDVLKPMKRLGALRSQIFGSIASASVFNLSYLPSTFTFWLPKPTMT EVSMTIVNRPDDIKDETAKVQTFGPGSTGFENNDSNHEHSTIINKGVYTRQGR LHVCGGISRKDDIAELPGIVHRQFADSNQREFRHRWHLDAESWDDVPDRSRNELK GRTRDLFGQLKSL"	
	complement (1991..2770)	
	/gene="thnF"	
gene	complement (1991..2770)	
	/gene="thnF"	
	/note="class II aldolase"	
	/codon_start=1	
CDS	/transl_table=11	
	/product="2,4-dihydroxydec-2-ene-1,10-dioic acid aldolase"	
	/protein_id="AA26442.1"	
	/db_xref="GI:23330203"	
gene	/translations="MENPONLQALLQKROIGLWLTIGDATSAEIICAAAGFDWLLI DGHVHSLDSVLAQRAISGYPTQAVRVSADPVIIKOYLDMGVQLLVPVQSG EBAAIIVRACRYPPGGERGVGGARRWGLIPDYLKADEQIAIIVQVTRQALDD AIMGTEGVGVFI GPADLSASLGHLCQPAHPDVARPEILHVIDRITRAGKAPGILTRDP EMLRBATERGALLNAVGLDASLLAAQAQTKRWKLMPSAERLT"	
	complement (1991..2770)	
	/gene="thnF"	
	/note="class II aldolase"	
CDS	/codon_start=1	
	/transl_table=11	
	/product="2,4-dihydroxydec-2-ene-1,10-dioic acid hydratase"	
	/protein_id="AA26441.2"	
gene	/db_xref="GI:23330202"	
	/translations="MLDDATIRBELAELHEARSGVQTKQSFQRPDIAFESRYVRC EMWIKLABGRRKIGKIGLTSRAMOKIANITPDYGVLLDDMMVADGSGVAPGRFIQ PFVEVELAFVINKDLGPDVSI FDI VLGATGTEWVVAVEIIDARIEDPTGSGRRVLD TIADNAAGIVVGVSSPVRPDALDLRWGVVTLTKNSVVEETGVGAGVLNHPAKGI VML ARKLSAMDEYLAAGELVLAGSFTRIVFVESGDQICADYGPLGTINVRFE"	
	complement (3584..4438)	
	/gene="thnD"	
CDS	complement (3584..4438)	
	/gene="thnD"	
	/codon_start=1	
	/transl_table=11	
gene	/product="2-oxodec-4-ene-1,10-dioic acid hydratase"	
	/protein_id="AA26440.1"	
	/db_xref="GI:10567587"	
	/translations="MSEAFVEKELPTKPYETHVMIGD PANPPLLVHVGARPGGSAGS NMBHCPADLYKDYVIAHDMTQGOSEIDPPVPAHAYTWGRIEQAQGVLDGLGIEK THIVGNSGGFLALHMLMEQDFDKVWTNGPAGAKMDATPELVLRVINYADPRLPRY REVINSFYDPSLVNLEEITQERFKIADDPVRRVQVEVMSEMTKMGMPALVVPVPSAL KRIPHEVALVGRHDRVIPLOASLYLLEHLQRAELFVLDRCGHWAQIQRDWAMLPIIR NHLELTAR"	
CDS	complement (4496..5308)	
	/gene="thnB"	
	complement (4496..5308)	
	/gene="thnB"	
gene	/codon_start=1	
	/transl_table=11	
	/product="cis-dihydrodiol dehydrogenase"	
	/protein_id="AA26445.1"	
CDS	/db_xref="GI:23330206"	
	/translations="MGWLDGKVALITGGGAGIGRALTVERFAEGAKVGVLEVNAGRAE ELAAVLGDSVCVVTGVSASPADNAAVQRTVEKFRGLDTFVANAGVDFDFPQAPQVEG DOLGAAPDELMNSVNVKGLLGVAALPELVSRGSGVITFISNAGFVAGGGPLYTASK HALVGLYQLAHELAPKYRVNAVAPGGMKTELSGLSATGTSQSLSQMDGLDGLDKDI TFQGVSPSPQICGPIVYLLASDSSVVTGAILINTDGGFVGRTIGVRGEGGL"	
	5702..6616	
	/gene="thnC"	
gene	5702..6616	
	/gene="thnC"	
	/codon_start=1	
	/transl_table=11	
CDS	/product="1,2-dihydroxynaphthalene dioxygenase ThnC"	
	/protein_id="AA252962.1"	
	/db_xref="GI:5817852"	
	/translations="MDKASVTELYGIVSVSDAAARVACCI VGMELVDDGEERFY LRWDGWHRIIVHTNGDDELYLGVNFDNLDATAAQLDAAGVSYSRVASAEAAER RVLLGLLEPQGNTEIFWGPQTRRPFHPQRMHGRFTGSOQLGCHLCRLRQDDDE AARVFTLFGKGVTEVHLPLPDGATAKPTFMHCNDRQHAVAFGLQMPKRNHLMIE VTELDLGIADHMDIREROIPVAMQLGKHSNDQALTFYHSPSGWLWELGWGVGKSSDQ LVVYKADLFGHGIEASGFLDVELGDDG"	
gene	5722..7036	
	/gene="thnA3"	
	5722..7036	
	/gene="thnA3"	
CDS	/codon_start=1	
	/transl_table=11	
	/product="Rieske-type ferredoxin"	
	/protein_id="AA252963.2"	
gene	/db_xref="GI:23330201"	
	/translations="MGRKVLCSASEVFPCEMLSVUTDAGDMLAVNVNVDGDFVTSNL	
	9.3%; Score 91.6; DB 1; Length 8154;	
	Best Local Similarity 49.7%; Pred. No. 1.5e-13;	
Query Match	Matches 316; Conservative	
	0; Mismatches 314; Indels 6; Gaps 3;	

791	QY	CAGTTACGGAAGACACCGNATAGCCTTTGGGAAGGCCCACTGGTTTTATTATCATCAGTTG	850
9609	Db	ACGTTTGAAGCAGCGCGGAAGCAGCAGGATGGAATGGTGATGTTGGGTTTTATCCATGATTTGT	9668
851	QY	TGCAGCCAGCGCTGCTTTGAAACCTACCGGAATACGAAATTTATCTTGC CGGTCCACCGC	910
9669	Db	TAGGAAAAACATTGGGAGAAAAACTTCCAGACCGAAATTTATTTCTGTGGCCCTCCCTC	9728
911	QY	CTATGCTCGACGCTACTGTCCGCTATGCTGCT --- CGGCAAGGGTGTTCCACGCGATCAA	967
9729	Db	CTATGACAGATGCTTTAAACAGGTATGCTGATGACTGAATACAAAGTGCCATTTGATCAA	9788
968	QY	TTCAATTTGACGCATTTTCTTAA	990
9789	Db	TTCAATTACGATCGTTTTTATTA	9811

Search completed: June 27, 2004, 07:56:13
Job time : 4017.03 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: June 27, 2004, 03:11:33 ; Search time 427.15 Seconds
(without alignments)
9846.004 Million cell updates/sec

Title: US-10-658-691-3
Perfect score: 990
Sequence: 1 atgtaccactcaaatga.....atttgacgcatattttctaa 990

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 3373863 seqs, 2124099041 residues

Total number of hits satisfying chosen parameters: 6747726

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : N Geneseqn_29Jan04:*
1: Geneseqn1980s:*
2: Geneseqn1990s:*
3: Geneseqn2000s:*
4: Geneseqn2001as:*
5: Geneseqn2001bs:*
6: Geneseqn2002s:*
7: Geneseqn2003as:*
8: Geneseqn2003bs:*
9: Geneseqn2003cs:*
10: Geneseqn2004s:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	990	100.0	990	2	AAV68080 DNA encod
2	117.2	11.8	5337	3	AAAI0957 DNA seque
3	107.8	10.9	5743	3	AAC63945 Burkholde
4	64.6	6.5	1180	4	AAC92962 Pseudomon
5	54.6	5.5	12808	3	AAAG5347 NDO relat
6	53	5.4	9706	3	AAAG5342 NDO relat
7	49.8	5.0	4355	3	AAAG5344 NDO relat
8	45.6	4.6	14462	3	AAAG5346 NDO relat
9	43	4.3	2000	7	ADA71938 Rice Gene
10	40.4	4.1	2000	7	ADA71938 Rice Gene
11	39	3.9	4721	3	AAAI1712 Comamonas
12	38.4	3.9	7785	4	ABL17653 Drosophil
13	38.4	3.9	12118	4	ABL17652 Drosophil
14	37.6	3.8	680	4	ABL15161 Drosophil
15	37.6	3.8	2738	4	ABL15160 Drosophil
16	37.6	3.8	6763	4	ABL07112 Drosophil
17	37	3.7	1029	7	ACA25179 Prokaryot
18	36.8	3.7	6379	2	AAQ79569 Nocardia
19	36.8	3.7	6379	2	AAQ79569 Nocardia
20	36.4	3.7	2394	7	ACA45674 Prokaryot
21	36.4	3.7	7188	7	ACC83205 Methane o
22	35.8	3.6	5331	3	AAA30292 DNA fragm
23	35.6	3.6	632	4	AAI23327 Probe #13

C 24	35.6	3.6	632	4	ABA68435 Human foe
C 25	35.6	3.6	632	4	AAI48649 Probe #17
C 26	35.6	3.6	632	4	ABA50485 Human bre
C 27	35.6	3.6	632	4	ABA35422 Probe #13
C 28	35.6	3.6	632	4	AAK42577 Human bon
C 29	35.6	3.6	632	4	AAK18808 Human bra
C 30	35.6	3.6	632	4	ABS42191 Human liv
C 31	35.6	3.6	632	5	AAI08968 Probe #89
C 32	35.6	3.6	632	6	ABS16628 Human gen
C 33	35.6	3.6	945	7	ACA25001 Prokaryot
C 34	35.6	3.6	1983	4	AAI14117 Probe #40
C 35	35.6	3.6	1983	4	ABA55843 Human foe
C 36	35.6	3.6	1983	4	AAI35499 Probe #41
C 37	35.6	3.6	1983	4	ABA45354 Human bre
C 38	35.6	3.6	1983	4	ABA25515 Probe #39
C 39	35.6	3.6	1983	4	AAK29543 Human bon
C 40	35.6	3.6	1983	4	AAK04063 Human bra
C 41	35.6	3.6	1983	4	ABS29170 Human liv
C 42	35.6	3.6	1983	5	AAI03970 Probe #39
C 43	35.6	3.6	1983	6	ABS04097 Human gen
C 44	34.8	3.5	65042	7	ABS55711 DNA encod
C 45	34.8	3.5	65042	8	ACA63029 Human gen

ALIGNMENTS

RESULT 1
AAV68080
ID AAV68080 standard; DNA; 990 BP.
XX
AC AAV68080;
XX
DT 12-JAN-1999 (first entry)
XX
DE DNA encoding an oxidase.
XX
KW Oxidase; dioxin removal; ds.
XX
OS Pseudomonas sp.
XX
PH Key Location/Qualifiers
FT CDS 1..990
FT /*tag= a
FT /product= "oxidase"
XX
PN JP10257895-A.
XX
PD 29-SEP-1998.
XX
PF 18-MAR-1997; 97JP-00084401.
XX
PR 18-MAR-1997; 97JP-00084401.
XX
PA (ASAH) ASAH KASEI KOGYO KK.
XX
DR WPI; 1998-575906/49.
DR P-PSDB; AAW80338.
XX
PT Oxidase gene derived from a microbe for removal of dioxin - converts hetero:poly:cyclic aromatic hydrocarbon to a benzoic acid homolog.
XX
PS Claim 2; Page 12-13; 15pp; Japanese.
XX
CC The present sequence encodes an oxidase which is derived from a microbe and converts a heteropolycyclic aromatic hydrocarbon to a benzoic acid homologue. The oxidase is used in a method for the removal of dioxin in which dibenzo-p-dioxin is converted to a diphenyl ether triol

XX Sequence 990 BP; 230 A; 248 C; 279 G; 233 T; 0 U; 0 Other;
SQ
Query Match 100.0%; Score 990; DB 2; Length 990;
Best Local Similarity 100.0%; Pred. No. 0;


```
QY 478 GGCATATGGGAGTCTTATTAAGAGGTACCCACAGCATTTAGTCTGCTTTTC 537
Db 4695 GGAGAGTGGCATTTCCAGATCCGGGGTGCCCAATGCAAGGAAGAGCATCTCGGCC 4754
QY 538 GAAATAGAAAGAGCGCTCGTCTATTTTGGACGGACCAATGGGCACATCTTTCTTC 597
Db 4755 GGGCATCTCAAAACAGGGGATTCGGTCAGGTTGAGGGCCCTATGGCCCTATTG 4814
QY 598 CGTCCAGGACGGCCGAAAGAGTCTTTGCAATTGGCGGGCTGCGGGCTCTGATGCG 657
Db 4815 AGAACCGAAATCGCGGGGACATCGCTGCGTGGCGGAGGCTCGGGGCTCGCGCCGTG 4874
QY 658 GCCGCTATTGCACGCG 673
Db 4875 GTGTCGATCGCGCGTG 4890

RESULT 3
AAC63945
ID AAC63945 standard; DNA; 5743 BP.
XX
AC AAC63945;
XX
DT 15-SEP-2003 (revised)
DI 09-FEB-2001 (first entry)
XX
DE Burkholderia sp. C1 desulphurase gene cluster.
XX
KW Desulphurase gene cluster; DBT dioxigenase; ferredoxin subunit;
KW Fe-S alpha subunit; dihydrodiol DBT dehydrogenase;
KW cis-3-hydroxythianaphenyl 2-oxo-3-butenate isomerase;
XX fossil fuel cleanup; ds.
OS Burkholderia sp; C1.
XX
FH Key Location/Qualifiers
CDS 78..1106
FT /*tag= a
FT /product= "DBT dioxigenase ferredoxin subunit (AAB29472)"
FT 2499..3332
FT /*tag= b
FT /product= "Dihydrodiol DBT dehydrogenase (AAB29473)"
FT 3389..4747
FT /*tag= c
FT /product= "DBT dioxigenase Fe-S alpha subunit (AAB29474)"
FT 4777..5274
FT /*tag= d
FT /product= "DBT dioxigenase Fe-S alpha subunit (AAB29475)"
FT 5455..5745
FT /*tag= e
FT /product= "cis-3-hydroxy-thianaphenyl 2-oxo-3-butenate
FT isomerase (AAB29476)"
XX
PN JP2000245477-A.
XX
PD 12-SEP-2000.
XX
PF 04-MAR-1999; 95JP-00057446.
XX
PR 04-MAR-1999; 95JP-00057446.
XX
PA (SEKI-) ZH SEKIYU SANGYO KASSEBIKA CENTER.
XX
WPI; 2000-622120/60.
DR P-PSDB; AAB29472, AAB29473, AAB29474, AAB29475, AAB29476.
XX
XX Gene encoding ferredoxin reductase subunit of DBT dioxigenase for
PT removing sulfur components contained in fossil fuel.
XX
PS Claim 1; Page 9-15; 24pp; Japanese.
XX
XX The invention relates to proteins from Burkholderia sp. C1 which have
```

```
CC desulphurase activity (AAB29472-B29476), and to the genes encoding them
CC (desulphurase gene cluster shown in AAC63945). The proteins of the
CC invention are a ferredoxin reductase subunit of DBT dioxigenase
CC (AAB29472), a dihydrodiol DBT dehydrogenase (AAB29473), Fe-S alpha
CC subunits of DBT dioxigenase (AAB29474, AAB29475) and a cis-3-hydroxy-
CC thianaphenyl 2-oxo-3-butenate isomerase (AAB29476). The invention also
CC relates to variants of the desulphurase proteins which retain activity,
CC and to constructs and transformants comprising DNA encoding the
CC desulphurase proteins. The desulphurase proteins can be used for removing
CC sulphur compounds from fossil fuels. The present sequence represents the
CC Burkholderia sp. C1 desulphurase gene cluster. (Updated on 15-SEP-2003 to
CC standardise OS field)
XX
```

QY Sequence 5743 BP; 1240 A; 1473 C; 1648 G; 1382 T; 0 U; 0 Other;

Query Match 10.9%; Score 107.8; DB 3; Length 5743;

Best Local Similarity 49.8%; Pred. No. 2.2e-24;

Matches 330; Conservative 0; Mismatches 327; Indels 6; Gaps 2;

```
QY 40 TCGGCTCAGGAAGAGCCTGTGTGCTTCAGCACCTTGTAAATGGTATCGAATTCGGTAC 99
Db 135 TGTTCACATGATGAACGCTACTCGCTCGGGGCTCAAGGCTGGCTTGGATTCCCTAT 194
QY 100 GAGTGTGCATCGGAGGTTGCGGAGTATGCAAAATTCGAGTTACTCGAAGGGGATGTCGA 159
Db 195 GAATGTCAATCCGATCGGTGCGAGCATTTGCCGCTTTCAGTTCTCGAGGGCGATGTCAAG 254
QY 160 TCAATGTGGCCGAGTGTCTCCAGGACTTTCTTCGCGAGATCGTGAGAAGGGCAACGCCAT 219
Db 255 GATTTGTGTCGAATGCGCGGGTTGTAATGCGAAGCTCGTGAATGCGGCA--TGCAC 311
QY 220 CTTGCATGCCAGTGGTTCGGCTCTCAGACCTTCGGGATCAAAAGTCGAGTCGAGGACAG 279
Db 312 CTCGGCTGCCAGACACACAGGTAGTACTCTCGAATCAAGCTGCGCTTGAAGCGGGC 371
QY 280 TAGCTCCCAACGATTCCCAATCTCAAGATGGAAGCGGAAGTTGTTGAGTCCGGCGCTA 339
Db 372 TATGTCGCCAGGTGATGCCGAGCGCTCAGCATGCGGCGTTCGTGATGTCGTGCCACTG 431
QY 340 ACTCATGACCTGCTGTCGTGCGATTAAGCATGATGGGCCAGCAAAATTTCTCCCGGCG 399
Db 432 AGCAAGACATGTTCCAGTTTCAGATTCCTACGAGGCTCCAGAGCTGAGTTTTGCTGGC 491
QY 400 CAGTTCTGCTAGTAGAGGCAGAGCATTTGCCAGCGTGTTCGCGCATATTCAATGGCG 459
Db 492 CAGTATGCCCTGCT---TCGCTTTCGCTGGCTGTCGGCACAAGGGCGTACTCGATGAGC 548
QY 460 AATTTAAAGAACCCCGAAGGCATATGGGAGTTCTATATTAAGAGGGTACCCACAGACGA 519
Db 549 AACCTGCCCAATGGCATTTGGCAATGGCAATTCATCGTCAACGCAAAACCGGCTGTTGT 608
QY 520 TTAGTCTCTGGCTTTTCGAAATAGAAAGAGGGCTCGTCTATTTTGGACGGGACCA 579
Db 609 GGAACCGCGGTGCTTTTCGATGTCCTGAAGAGGTTGATGCTATTGAGTTGAGGGTCTCT 668
QY 580 ATGGGCACATCTTTCTTCCTCCAGGGACCGGCCGGAAGAGTCTTTTGCATTTGGCGCGGT 639
Db 669 TACGGACCGCTTATCTGCTACCGAAACCGGTAGAGGTGTCGTCTGCAATTTGTCGCGC 728
QY 640 GCGGGCTCTCGTATGCGGCGCGCTATTGCAACCGGCTCGATCGCGGAAACACAGACCG 599
Db 729 TCGGACTTTTCCGATGTTGTCAATCTCTGCGCGCGAGTGTGCAATCCGCAATGACC 788
QY 700 GTA 702
Db 789 GAA 791
```

RESULT 4

AAC92962

ID AAC92962 standard; DNA; 1180 BP.

XX

AC AAC92962;

XX 27-MAR-2001 (first entry)
 XX Pseudomonas mendocina KRI tmo F gene.
 XX Monooxygenase; epoxide preparation; enantio-specific;
 KW non-haem diiron-containing monooxygenase; alkene epoxidation; bioreactor;
 KW pharmaceutical production; toluene monooxygenase;
 KW toluene-4-monooxygenase; T4MO; tmo F; ds.
 XX Pseudomonas mendocina.
 OS WO200073425-A1.
 XX 07-DEC-2000.
 XX 26-MAY-2000; 2000WO-US014537.
 XX 28-MAY-1999; 99US-0136602P.
 XX (ENVI-) ENVIROGEN INC.
 XX Steffan RJ, Mcclay KR;
 PI WPI: 2001-049935/06.
 DR P-PSDB; AAB60228.
 XX Preparation of enantio-specific epoxides useful in synthetic organic
 PT chemistry and pharmaceutical reactions, by contacting alkene with an
 PT enzyme comprising a native or mutated non-haem diiron-containing
 PT monooxygenase.
 XX Example 3; Fig 16B; 70pp; English.
 XX The invention relates to a method for the enantio-specific preparation of
 CC epoxides using a wild-type or a mutant non-haem diiron-containing
 CC monooxygenase, and an alkene substrate. In particular, the non-haem
 CC diiron-containing monooxygenase is a toluene monooxygenase. The invention
 CC also relates to mutant non-haem diiron-containing monooxygenases which
 CC produce a different ratio of (R) and (S) enantiomers of an epoxide
 CC relative to the wild-type enzyme, and methods of producing such mutant
 CC enzymes. The method is used for large scale production of epoxides which
 CC are useful in many processes, particularly in synthetic organic chemistry
 CC and in the production of pharmaceuticals. In an exemplary approach,
 CC alkenes are passed through a fluid-bed reactor inoculated with bacteria
 CC such as Pseudomonas mendocina KRI possessing or transformed with a
 CC mutated non-haem diiron-containing monooxygenase. The present sequence
 CC represents the Pseudomonas mendocina KRI toluene-4-monooxygenase (T4MO)
 CC tmo F subunit gene
 XX Sequence 1180 BP; 300 A; 256 C; 311 G; 313 T; 0 U; 0 Other;
 SQ Query Match 5.5%; Score 64.6; DB 4; Length 1180;
 Best Local Similarity 51.0%; Pred. No. 2.5e-10;
 Matches 179; Conservative 0; Mismatches 169; Indels 3; Gaps 1;
 QY 68 CAGCAGTGTGTAATGGTATCGGATTTCCGACAGAGTGTGCATCGGAGGTTGGGAGTAT 127
 DB 119 CTGCTACGTGCTGAATTGGTATTTCCATATGATGTTACTCAGAGGTTGGGCGCAT 178
 QY 128 GCATATTCGAGTACTCGAAGGATGTCCAAATCAATGTGCGCGATGCTCCAGGACTTT 187
 DB 179 GTAAGATCGAGTGTCTTGAGGGAGAGTCTCTAACTATGCGCTGATGCACAGGATTAG 238
 QY 188 CTTTCGCGAGATCGTGAGAAGGGAACCGCATCTTGTGATCCGAGTGGTTCGCTCTCAG 247
 DB 239 CCGCCCTGACTCCGTAAG--AATCGTTTTTGGCGTCCAGTGCACACATTATCCG 295
 QY 248 ACCTCGCGATCAAGTCCGAGTGCAGGACGAAGTACGTCCCAAGATTCCAAATCTCAAGAA 307
 DB 296 ACCTCAAAATTAAGGTATTAAACCGTCCGGAGGAGCGTGTTCACATCCCGCCCAACGTT 355
 QY 308 TGGAAAGCGGAAGTTGTTGAGTCCGGCGGCTAACTCATGACCTGCTCGTCCGATTAC 367

DB 356 TCTCGACTCGAGTAGTTAGTAAGCGCTTCTCTCTGACGAGATGTTGAGTCGGACTTG 415
 QY 368 GCACCTGATGGCCAGCAAAATTTCTCCCGCCAGTTCCTGCTAGTAGTAGG 418
 DB 416 AAGCGGACAGAAAGTGTGTTTTCACAGGCGCATATTTATGTTGACG 466
 RESULT 5
 AAA65347
 ID AAA65347 standard; DNA; 12808 BP.
 XX AC AAA65347;
 XX 09-NOV-2000 (first entry)
 XX NDO related complex alpha subunit nucleotide sequence SEQ ID NO:10.
 KW Pseudomonas sp. strain NCIB 9816-4; naphthalene dioxygenase; NDO;
 KW inducible multi-component enzyme; alpha subunit; mutant; chiral diol;
 KW polymer; resin; pharmaceutical; rubber industry; bioremediation; ds.
 XX Pseudomonas sp.
 OS Synthetic.
 XX WO200037480-A1.
 XX 29-JUN-2000.
 XX 26-OCT-1999; 99WO-US025079.
 XX 26-OCT-1998; 98US-0105575P.
 XX (IOWA) UNIV IOWA RES FOUND.
 XX Parales R, Gibson D, Resnick S, Lee K;
 XX WPI: 2000-452174/39.
 DR P-PSDB; AAB12573.
 XX Novel naphthalene dioxygenase mutant having a specific amino acid
 PT substitution for preparing chiral diols for use in the polymer, resin,
 PT pharmaceutical or rubber industry and for carrying out bioremediation.
 XX Disclosure; Page 81-86; 151pp; English.
 XX The present invention describes a naphthalene dioxygenase (NDO) or NDO
 CC related complex (I) comprising several polypeptides which contain an
 CC alpha subunit that contains substituted amino acids at specific
 CC positions. The polypeptides and host cells are useful for preparing (-)-
 CC (1S,2R)-cis-naphthalene dihydrodiol, (- or +)-cis-biphenyl-3,4-
 CC dihydrodiol, (1S,2R)-cis-phenanthrene-1,2-dihydrodiol, cis-1,2-
 CC dihydroxyindan, 1,2-dihydroxy-1,2,3,4-tetrahydronaphthalene, which
 CC involves contacting them with naphthalene, biphenyl, phenanthrene,
 CC indene, and 1,2-dihydronaphthalene respectively. The polypeptides and the
 CC host cells are also useful for preparing 1,2-dihydroxy-1,2-
 CC dihydrophenanthrene or 3,4-dihydroxy-3,4-dihydrophenanthrene which
 CC involves contacting them with phenanthrene. The polypeptides and the host
 CC cells are also used in bioremediation in which they oxidize an aromatic
 CC compound such as indene, 1,2-dihydronaphthalene, benzocyclopent-1-ene,
 CC anthracene, phenanthrene, dibenzo(1,4)dioxan, acenaphthylene,
 CC naphthalene, biphenyl, fluorene, dibenzofuran, benzo[thiophene], 9,10-
 CC dihydroanthracene, or 9,10-dihydrophenanthrene a corresponding dihydro
 CC dihydroxy compound. The polypeptides and the host cells are useful for
 CC preparing chiral diols for use in the polymer, resin, pharmaceutical or
 CC rubber industry. The present sequence encodes a naphthalene dioxygenase
 CC (NDO) mutant alpha subunit, given in the exemplification of the present
 CC invention
 SQ Sequence 12808 BP; 2992 A; 3328 C; 3576 G; 2912 T; 0 U; 0 Other;
 Query Match 5.5%; Score 54.6; DB 3; Length 12808;
 Best Local Similarity 50.4%; Pred. No. 2e-06;

	Matches	191; Conservative	0; Mismatches	179; Indels	9; Gaps	2;
Qy	303	AAGAATGGAGACGGGAAGTTGTTGAGGTCCGGGGCTAACTCATGACCTGCTGCTCCGGTCCG	362			
Db	1101	AATCATCAAAAGGCATGTTGTGCGGTGAGTCGCCACATCAGATATTCGCGCATCCG	1160			
Qy	363	ATTACGCATGATGGGCACGAATTTCTCTCCCGGCCAGTTCTGCTCTAGTAGGGCAGA	422			
Db	1161	CATTCCGCTCGCCAAAGCCCTTCGAGTTCTCACCGGACAGTAC-----GCGATGCTACA	1214			
Qy	423	GCAGTTGCCAGGCGTGTTCGCGCATATTCAATGGCGAATTTAAAGAACCCCGAAGGCAT	482			
Db	1215	GTTTCAGTCCCGACATGTGGCTGCATATTTCAATGGCTGTCTGCCAGATGACCAAGAAAT	1274			
Qy	483	ATCGGAGTTCTATATTAAAGAGGTACCCACAGACGATTTAGTCTCTTGGCTTTTCGAAA	542			
Db	1275	--GGAATTCCATATCCGCAAGTCCGGGCGCGGTGTCAAGAGTATATTTTCGAGCA	1331			
Qy	543	TAGAAAAGRAGCGCTCGTCTATTTTTCGCGGACCAATCGGCACATCTTCTTCGGTCC	602			
Db	1332	CGTCCGGAAGGTACAAGCATTAAGTTGATGGGCTCTTGGTACGGCCCTATCTGGGTCA	1391			
Qy	603	AGG3ACCGGCGCGAAAGAGTCTTTGCATTTGGCGCGGTCCCGGGCTCTCGTATGCGGGCCG	662			
Db	1392	GGCTCACACGGGCCGATGTGTGTGTGGCGCGGACCGGACTCGACCGGTGCTGTC	1451			
Qy	663	TATTGCACGCGCTCGATG	681			
Db	1452	GATTGTTCCGGCGCGCTG	1470			

RESULT 6
 AAA65342
 ID AAA65342 standard; DNA; 9706 BP.
 XX
 AC AAA65342;
 XX
 DT 09-NOV-2000 (first entry)
 XX
 DE NDO related complex alpha subunit nucleotide sequence SEQ ID NO:5.
 XX
 KW Pseudomonas sp. strain NCIB 9816-4; naphthalene dioxygenase; NDO;
 KW inducible multi-component enzyme; alpha subunit; mutant; chiral diol;
 KW polymer; resin; pharmaceutical; rubber industry; bioremediation; ds.
 XX
 OS Pseudomonas sp.
 OS Synthetic.
 XX
 PN WO200037480-A1.
 XX
 PD 29-JUN-2000.
 XX
 XX 26-OCT-1999; 99WO-US025079.
 XX
 PR 26-OCT-1998; 98US-0105575P.
 XX
 PA (IOWA) UNIV IOWA RES FOUND.
 XX
 PI Parales R, Gibson D, Resnick S, Lee K;
 XX
 DR WPI; 2000-452174/39.
 DR P-PSDB; AAB12568.
 XX
 PT Novel naphthalene dioxygenase mutant having a specific amino acid
 PT substitution for preparing chiral diols for use in the polymer, resin,
 PT pharmaceutical or rubber industry and for carrying out bioremediation.
 XX
 PS Disclosure; Page 66-70; 151pp; English.
 XX
 CC The present invention describes a naphthalene dioxygenase (NDO) or NDO
 CC related complex (I) comprising several polypeptides which contain an
 CC alpha subunit that contains substituted amino acids at specific
 CC positions. The polypeptides and host cells are useful for preparing (-)-

CC	(1S,2R)-cis-naphthalene dihydrodiol, (- or +)-cis-biphenyl-3,4-dihydrodiol, (1S,2R)-cis-phenanthrene-1,2-dihydrodiol, cis-1,2-dihydroxyindan, 1,2-dihydroxy-1,2,3,4-tetrahydronaphthalene, which involves contacting them with naphthalene, biphenyl, phenanthrene, indene, and 1,2-dihydronaphthalene respectively. The polypeptides and the host cells are also useful for preparing 1,2-dihydroxy-1,2-dihydronaphthalene or 3,4-dihydroxy-3,4-dihydronaphthalene which involves contacting them with phenanthrene. The polypeptides and the host cells are also used in bioremediation in which they oxidize an aromatic compound such as indene, 1,2-dihydronaphthalene, benzocyclohept-1-ene, anthracene, phenanthrene, dibenzo(1,4)dioxan, acenaphthylene, naphthalene, biphenyl, fluorene, dibenzofuran, dibenzothiophene, 9,10-dihydroanthracene, or 9,10-dihydronaphthalene a corresponding dihydroxy compound. The polypeptides and the host cells are useful for preparing chiral diols for use in the polymer, resin, pharmaceutical or rubber industry. The present sequence encodes a naphthalene dioxygenase (NDO) mutant alpha subunit, given in the exemplification of the present CC invention
XX	
SQ	Sequence 9706 BP; 2331 A; 2528 C; 2629 G; 2218 T; 0 U; 0 Other;
	Query Match 5.4%; Score 53; DB 3; Length 9706;
	Best Local Similarity 50.1%; Pred. No. 6e-06;
	Matches 190; Conservative 0; Mismatches 180; Indels 9; Gaps 2;
QY	303 AAGAAATCGAAGCGGAAGTTGCTTGAAGTCGGGGCGCTAACTCATGACCTGTGTCTCGTGCG 362
Db	1193 AATCATCAAGGCACAGTAGTGTCTCAGTTCGAGTCGCCCACTCACGATATCCGTTCGCTTACG 1252
QY	363 ATTACGCACCTGATTGGCCAGCAAATTTCTCCCCGGCCAGTTCTGCCTAGTAGAGGCAGA 422
Db	1253 CGTACGCCTCTCCAAGCCCTTCGAGTTCTCACCCGGACAGTACGCGACACT-----GCA 1306
QY	423 GCAGTTGCCAGGCGTGGTTCGGCCATATTCATATGGCGAATTTAAAGAACCCCAGAGCAT 482
Db	1307 GTTCAGCCCTTGACATCGCGCGTCCGCTATTCAATTGGCAGGTTTCCAGATGACCAGAAAT 1366
QY	483 ATCGGAGTTCTATATAAGAGGGTACCCACAGAACGATTTAGTCCCTTGGCTTTTCGAAAA 542
Db	1367 ---GGAGTTCCACATACGCGAGTTCGCGGTGGCGCGTCACGGAGTAGTTTTCGAACA 1423
QY	543 TAGAAAAGAAGCGCTCGTCTATTTTTGACGGGACCAATGGGGCACATCTTTCTTCGGTCC 602
Db	1424 CGTCCGCGAAGGTACAAAGCATCAAGTTGAGCGGGCCCTCTTGGTACGGCTTATCTACGTCA 1483
QY	603 AGGGACCGGCGGAAAAGAGTCTTTTGCAATGTCGGCGGTGCGGGCTCTCGTATCGGGCCG 662
Db	1484 GAAGCACACCGGACCGATGCTGTGTGTAGTGTSCGGGACCGGACTGCACCGGTGCTGTC 1543
QY	663 TATTGCAOCCGCTCGATG 681
Db	1544 GATTGTTCCGGCGCGCTG 1562

RESULT 7	
AAA65344	
AAA65344	
ID	AAA65344 standard; DNA; 4355 BP.
XX	
XX	
AC	AAA65344;
XX	
XX	
DT	C9-NOV-2000 (first entry)
XX	
DE	NDO related complex alpha subunit nucleotide sequence SEQ ID NO:7.
XX	
XX	
KW	<i>Pseudomonas</i> sp. strain NCIB 9816-4; naphthalene dioxygenase; NDO;
KW	inducible multi-component enzyme; alpha subunit; mutant; chiral diol;
KW	polymer; resin; pharmaceutical; rubber industry; bioremediation; ds.
XX	
XX	
OS	<i>Pseudomonas</i> sp.
XX	
OS	Synthetic.
XX	
PN	WQ200037480-A1.
XX	
XX	

29-JUN-2000.
26-OCT-1999; 99WO-US025079.
26-OCT-1998; 98US-0105575P.
(IOWA) UNIV IOWA RES FOUND.
Parales R, Gibson D, Resnick S, Lee K;
WPI; 2000-452174/39.
P-PSDB; AAB12570.
Novel naphthalene dioxygenase mutant having a specific amino acid substitution for preparing chiral diols for use in the polymer, resin, pharmaceutical or rubber industry and for carrying out bioremediation.
Disclosure; Page 71-73; 151pp; English.
The present invention describes a naphthalene dioxygenase (NDO) or NDO related complex (I) comprising several polypeptides which contain an alpha subunit that contains substituted amino acids at specific positions. The polypeptides and host cells are useful for preparing (-)-(1S,2R)-cis-naphthalene dihydrodiol, (- or +)-cis-biphenyl-3,4-dihydrodiol, (1S,2R)-cis-phenanthrene-1,2-dihydrodiol, cis-1,2-dihydroxyindan, 1,2-dihydroxy-1,2,3,4-tetrahydronaphthalene, which involves contacting them with naphthalene, biphenyl, phenanthrene, indene, and 1,2-dihydronaphthalene respectively. The polypeptides and the host cells are also useful for preparing 1,2-dihydroxy-1,2-dihydronaphthalene or 3,4-dihydroxy-3,4-dihydronaphthalene which involves contacting them with naphthalene, biphenyl, phenanthrene, indene, and 1,2-dihydronaphthalene respectively. The polypeptides and the host cells are also used in bioremediation in which they oxidize an aromatic compound such as indene, 1,2-dihydronaphthalene, benzocyclohept-1-ene, anthracene, phenanthrene, dibenzo(1,4)dioxan, acenaphthylene, naphthalene, biphenyl, fluorene, dibenzofuran, dibenzothiophene, 9,10-dihydroanthracene, or 9,10-dihydronaphthalene a corresponding dihydroxy compound. The polypeptides and the host cells are useful for preparing chiral diols for use in the polymer, resin, pharmaceutical or rubber industry. The present sequence encodes a naphthalene dioxygenase (NDO) mutant alpha subunit, given in the exemplification of the present invention
Sequence 4355 BP; 1065 A; 1143 C; 1171 G; 976 T; 0 U; 0 Other;
Query Match 5.0%; Score 49.8; DB 3; Length 4355;
Best Local Similarity 50.1%; Pred. No. 4.5e-05;
Matches 182; Conservative 0; Mismatches 172; Indels 9; Gaps 2;
QY 319 GTTGTGAGTCCGGCGCTAACTCATGACCTGCTCGTGGGATAGCGACTGATGGG 378
1159 GTGGTCGCGGTGAGTGGCGGCACTACGATATCCGTCGCTACGCGTACGCTCGCTAAG 1218
379 CCAGCAAAATTCCTCCCGGCCAGTTCGCTAGTAGAGCGCAGCGTGCAGCGGTG 438
1219 CCCTTCGAGTTCACCCGGACAGTAGCGACATT-----GCAGTTCAGTCTGAGCAT 1272
439 GTTCGGCGCATATTCAATGCGGAATTTAAAGACCCCGAAGGCGATATGGAGTTCATATT 498
1273 GCGCGTCCGTAATCAATGCGGAGTCTGCCAGATGACCAAGAAAT---GGAGTTCACATA 1329
499 AAGAGGGTACCCACAGACGATTTAGTCTTCGCTTTTCGAAATAGAAAGAGCGCT 558
1330 CGCAAGGTGCGGTGGCGCGTAACGGAGTAGTGTTCGACGACGTCGCGAAGGTACA 1389
559 CGTCTATTTTACGGGACCAATAGGCGACATCTTTCTTCGTCAGGGACCGCGCGAAG 618
1390 AGCATCAAGTTGAGCGGCGCACTTGGTAGCGCTTATTTGCGTCAGAACACACCGCGCG 1449
619 AGCTTTTGCATTTGGCGCGGTGCGGCGCTCTGTATGCGGCGCTATTTCAGCGCGCTCG 678
1450 ATGCTCTGTGTGGCGGTGGACCGGACTAGCACCGGTGCTGTCGATTGTTCCGCGCGG 1509
679 ATG 681

Db 1510 CTG 1512
RESULT 8
AAA65346
ID AAA65346 standard; DNA; 14462 BP.
XX AC AAA65346;
XX AC AAA65346;
DT 09-NOV-2000 (first entry)
XX NDO related complex alpha subunit nucleotide sequence SEQ ID NO:9.
DE Pseudomonas sp. strain NCIB 9816-4; naphthalene dioxygenase; NDO;
KW inducible multi-component enzyme; alpha subunit; mutant; chiral diol;
KW polymer; resin; pharmaceutical; rubber industry; bioremediation; ds.
XX Pseudomonas sp.
OS Synthetic.
OS WO200037480-A1.
XX 29-JUN-2000.
XX 26-OCT-1999; 99WO-US025079.
XX 26-OCT-1998; 98US-0105575P.
XX (IOWA) UNIV IOWA RES FOUND.
XX Parales R, Gibson D, Resnick S, Lee K;
XX WPI; 2000-452174/39.
XX P-PSDB; AAB12572.
XX Novel naphthalene dioxygenase mutant having a specific amino acid substitution for preparing chiral diols for use in the polymer, resin, pharmaceutical or rubber industry and for carrying out bioremediation.
XX Disclosure; Page 75-81; 151pp; English.
The present invention describes a naphthalene dioxygenase (NDO) or NDO related complex (I) comprising several polypeptides which contain an alpha subunit that contains substituted amino acids at specific positions. The polypeptides and host cells are useful for preparing (-)-(1S,2R)-cis-naphthalene dihydrodiol, (- or +)-cis-biphenyl-3,4-dihydrodiol, (1S,2R)-cis-phenanthrene-1,2-dihydrodiol, cis-1,2-dihydroxyindan, 1,2-dihydroxy-1,2,3,4-tetrahydronaphthalene, which involves contacting them with naphthalene, biphenyl, phenanthrene, indene, and 1,2-dihydronaphthalene respectively. The polypeptides and the host cells are also useful for preparing 1,2-dihydroxy-1,2-dihydronaphthalene or 3,4-dihydroxy-3,4-dihydronaphthalene which involves contacting them with naphthalene, biphenyl, phenanthrene, indene, and 1,2-dihydronaphthalene respectively. The polypeptides and the host cells are also used in bioremediation in which they oxidize an aromatic compound such as indene, 1,2-dihydronaphthalene, benzocyclohept-1-ene, anthracene, phenanthrene, dibenzo(1,4)dioxan, acenaphthylene, naphthalene, biphenyl, fluorene, dibenzofuran, dibenzothiophene, 9,10-dihydroanthracene, or 9,10-dihydronaphthalene a corresponding dihydroxy compound. The polypeptides and the host cells are useful for preparing chiral diols for use in the polymer, resin, pharmaceutical or rubber industry. The present sequence encodes a naphthalene dioxygenase (NDO) mutant alpha subunit, given in the exemplification of the present invention
Sequence 14462 BP; 3426 A; 3749 C; 4018 G; 3269 T; 0 U; 0 Other;
SQ Query Match 4.6%; Score 45.6; DB 3; Length 14462;
Best Local Similarity 52.0%; Pred. No. 0.0022;
Matches 102; Conservative 0; Mismatches 94; Indels 0; Gaps 0;
QY 486 GGAGTTCATATTAGAGGTACCCACAGGACGATTTAGTCTTCGTTTCGAAATAG 545

XX	Phenol hydroxylase; microbe; phenol decomposition; ds.
XX	Comamonas testosteroni.
XX	Key
XX	Location/Qualifiers
XX	19..210
XX	/tag= a
XX	/product= "protein 1"
XX	232..11244
XX	/tag= b
XX	/product= "protein 2"
XX	1256..1549
XX	/tag= c
XX	/product= "protein 3"
XX	1597..3219
XX	/tag= e
XX	/product= "protein 4"
XX	3216..3572
XX	/tag= f
XX	/product= "protein 5"
XX	3632..4705
XX	/tag= g
XX	/product= "protein 6"
XX	JP2000069968-A.
XX	PN
XX	07-MAR-2000.
XX	28-AUG-1998; 98JP-002432249.
XX	28-AUG-1998; 98JP-002432249.
XX	(KAIY-) KAIYO BIOTECHNOLOGY KENKYUSHO KK.
XX	WPI; 2000-264445/23.
XX	P-PSDB; AAW90942, AAW90943, AAW90944, AAW90945, AAW90946, AAW90947.
XX	Structural gene and a regulator gene of phenol hydroxylase of Comamonas
XX	testosteroni R5 - used for decomposing phenol.
XX	Disclosure; Page 6-11, 18pp; Japanese.
XX	This invention describes a novel microbe for decomposing phenol which
XX	carries a phenol hydroxylase protein. This sequence encodes the proteins
XX	AAW90942-W90947 which are described in the method of the invention.
XX	(Updated on 06-AUG-2003 to correct OS field.)
XX	Sequence 4721 BP; 908 A; 1548 C; 1424 G; 841 T; 0 U; 0 Other;
XX	Query Match 3.9%; Score 39; DB 3; Length 4721;
XX	Best/Local Similarity 52.8%; Pred. No. 0.19; Indels 0; Gaps 0;
XX	Matches 84; Conservative 0; Mismatches 75;
QY	783 CGTCCAGGCAGTTACGGAAGACACGCGATACGCTTTGCGCAAGGCCCCACGTGGTTATTCA 842
Db	4450 CATCCCGCGATATCGGCGAGACTTCGCGGGATTGCGAAGCGCGCTTGTCGA 4509
QY	843 TCAGGTTGTCGACGACGCTGCTTGAACCCCTACCGGAATACGAAATTTATCTTGC CGG 902
Db	4510 CGAGGCTGCGAAGCCATTTCGAAGCGGCTTTCGCGGTTGCAAGCCATCTCTGCGG 4569
QY	903 TCACCGCCTATGGTCGACGCTACTCTCCGATGCTGCT 941
Db	4570 GCCACCAACCATGTCGAAGCCTGTATTTCCACCCCTGAT 4608
XX	RESULT 12
XX	ABL17653
XX	ID ABL17653 standard; DNA; 7785 BP.
XX	AC ABL17653;
XX	XX

DT 26-MAR-2002 (first entry)
XX Drosophila melanogaster genomic polynucleotide SEQ ID NO 4432.
DE
XX
XX Drosophila melanogaster genomic polynucleotide; cell signalling; insecticide;
KW
XX Drosophila melanogaster; gene; ds.
KW
XX Drosophila melanogaster.
OS
XX
XX Drosophila melanogaster.
PN
XX WO200171042-A2.
XX
XX 27-SEP-2001.
XX
XX 23-MAR-2001; 2001WO-US009231.
XX
XX 23-MAR-2000; 2000US-0191637P.
PR
XX 11-JUL-2000; 2000US-00614150.
XX
XX (PEKE) PE CORP NY.
PA
XX
XX Venter JC, Adams M, Li PWD, Myers EW;
PI
XX WPI; 2001-656860/75.
DR
XX
XX New isolated nucleic acid detection reagent for detecting 1000 or more
PT genes from Drosophila and for elucidating cell signalling and cell-cell
PT interactions.
XX
XX Claim 1; SEQ ID NO 4432; 21pp + Sequence Listing; English.
XX
XX The invention relates to an isolated nucleic acid detection reagent
PS capable of detecting 1000 or more genes from Drosophila. The invention is
XX useful in developmental biology and in elucidating cell signalling and
XX cell-cell interactions in higher eukaryotes for the development of
XX insecticides, therapeutics and pharmaceutical drugs. The invention
XX discloses genomic DNA sequences (ABL16176-ABL30511), expressed DNA
XX sequences (ABL01840-ABL16175) and the encoded proteins (ABB57737-
XX ABB72072). The sequence data for this patent did not form part of the
XX printed specification, but was obtained in electronic format directly
XX from WIPO at ftp.wipo.int/pub/published_pct_sequences
XX
SQ Sequence 7785 BP; 2166 A; 2186 C; 2007 G; 1426 T; 0 U; 0 Other;
Query Match 3.9%; Score 38.4; DB 4; Length 7785;
Best Local Similarity 49.0%; Pred. No. 0.4;
Matches 102; Conservative 0; Mismatches 106; Indels 0; Gaps 0;
QY 127 TGCAAATTCGAGTTCTACTCGAAGGGAATGTCCAATCAATGTGGCGGATGCTCCAGGACTT 186
Db 7243 TGCAAATTCGAGTTCTACTCGAAGGGAATGTCCAATCAATGTGGCGGATGCTCCAGGACTT 7302
QY 187 TCTTCGGGAGATCGTGAGAGGGCAACCGGCATCTTGCATGCCAGTCCGCTGCTCA 246
Db 7303 GTGTCCACATCTGTGCCAAGCTCGGTCTCCATGCCCATCATCGCGCAACTGCTGTTTC 7362
QY 247 GACCTTGGGATCAAAAGTCGAGTGCAGACAGTACGTCCTCAAGATTCCTCAAGA 306
Db 7363 TACTTGAGGGATAAAATCCCTTCAGTTGCGATTCTCTCAAGGAGCAAAAGTCAAA 7422
QY 307 ATGGAAGCGGAAGTTGTTGAGTCCGGG 334
Db 7423 TTCGACACGGAACCCATCGAGATCAAG 7450
RESULT 13
ABL17652/c
ID ABL17652 standard; DNA; 12118 BP.
XX
XX ABL17652;
AC
XX 26-MAR-2002 (first entry)
DT
XX Drosophila melanogaster genomic polynucleotide SEQ ID NO 4429.
DE

XX Drosophila; developmental biology; cell signalling; insecticide;
KW pharmaceutical; gene; ds.
XX
XX Drosophila melanogaster.
OS
XX WO200171042-A2.
XX
XX 27-SEP-2001.
XX
XX 23-MAR-2001; 2001WO-US009231.
XX
XX 23-MAR-2000; 2000US-0191637P.
PR
XX 11-JUL-2000; 2000US-00614150.
XX
XX (PEKE) PE CORP NY.
PA
XX
XX Venter JC, Adams M, Li PWD, Myers EW;
PI
XX WPI; 2001-656860/75.
DR
XX
XX New isolated nucleic acid detection reagent for detecting 1000 or more
PT genes from Drosophila and for elucidating cell signalling and cell-cell
PT interactions.
XX
XX Claim 1; SEQ ID NO 4429; 21pp + Sequence Listing; English.
XX
XX The invention relates to an isolated nucleic acid detection reagent
PS capable of detecting 1000 or more genes from Drosophila. The invention is
XX useful in developmental biology and in elucidating cell signalling and
XX cell-cell interactions in higher eukaryotes for the development of
XX insecticides, therapeutics and pharmaceutical drugs. The invention
XX discloses genomic DNA sequences (ABL16176-ABL30511), expressed DNA
XX sequences (ABL01840-ABL16175) and the encoded proteins (ABB57737-
XX ABB72072). The sequence data for this patent did not form part of the
XX printed specification, but was obtained in electronic format directly
XX from WIPO at ftp.wipo.int/pub/published_pct_sequences
XX
SQ Sequence 12118 BP; 2692 A; 2753 C; 3140 G; 3533 T; 0 U; 0 Other;
Query Match 3.9%; Score 38.4; DB 4; Length 12118;
Best Local Similarity 49.0%; Pred. No. 0.51;
Matches 102; Conservative 0; Mismatches 106; Indels 0; Gaps 0;
QY 127 TGCAAATTCGAGTTCTACTCGAAGGGAATGTCCAATCAATGTGGCGGATGCTCCAGGACTT 186
Db 1612 TGCAAATTCGAGTTCTACTCGAAGGGAATGTCCAATCAATGTGGCGGATGCTCCAGGACTT 1553
QY 187 TCTTCGGGAGATCGTGAGAGGGCAACCGGCATCTTGCATGCCAGTCCGCTCTCA 246
Db 1552 GTGTCCACATCTGTGCCAAGCTCGGTCTCCATGCCCATCATCCGCAACTGCTGTTTC 1493
QY 247 GACCTTGGGATCAAAAGTCCGAGTGCAGACAGTACGTCCTCAAGATTCCTCAAGA 306
Db 1492 TACCTGAGGGATAAAATCCCTTCAGTTGCGATTCTCTCAAGGAGCAAAACGTCAAA 1433
QY 307 ATGGAAGCGGAAGTTGTTGAGTCCGGG 334
Db 1432 TTCGACACGGAACCCATCGAGATCAAG 1405
RESULT 14
ABL15161/c
ID ABL15161 standard; cDNA; 680 BP.
XX
XX ABL15161;
AC
XX 26-MAR-2002 (first entry)
DT
XX Drosophila melanogaster expressed polynucleotide SEQ ID NO 39965.
DE
XX Drosophila; developmental biology; cell signalling; insecticide;
KW pharmaceutical; gene; ss.

```
XX OS Drosophila melanogaster.
XX PN WO200171042-A2.
XX PD 27-SEP-2001.
XX XX
XX PF 23-MAR-2001; 2001WO-US009231.
XX PR 23-MAR-2000; 2000US-0191637P.
XX PR 11-JUL-2000; 2000US-00614150.
XX PA (PEXE ) PE CORP NY.
XX PI Venter JC, Adams M, Li PWD, Myers EW;
XX DR WPI; 2001-656860/75.
XX DR P-PSDB; ABB71058.
XX PT New isolated nucleic acid detection reagent for detecting 1000 or more
XX PT genes from Drosophila and for elucidating cell signaling and cell-cell
XX PT interactions.
XX PS Claim 1; SEQ ID NO 39965; 21pp + Sequence Listing; English.
XX CC The invention relates to an isolated nucleic acid detection reagent
XX CC capable of detecting 1000 or more genes from Drosophila. The invention is
XX CC useful in developmental biology and in elucidating cell signaling and
XX CC cell-cell interactions in higher eukaryotes for the development of
XX CC insecticides, therapeutics and pharmaceutical drugs. The invention
XX CC discloses genomic DNA sequences (ABL16176-ABL30511), expressed DNA
XX CC sequences (ABL01840-ABL16175) and the encoded proteins (ABB57737-
XX CC ABB72072). The sequence data for this patent did not form part of the
XX CC printed specification, but was obtained in electronic format directly
XX CC from WIPO at ftp.wipo.int/pub/published_pct_sequences
XX SQ Sequence 680 BP; 193 A; 163 C; 164 G; 160 T; 0 U; 0 Other;

Query Match          3.8%; Score 37.6; DB 4; Length 680;
Best Local Similarity 54.3%; Pred. No. 0.19;
Matches 76; Conservative 0; Mismatches 64; Indels 0; Gaps 0;

QY 751 GATATCGACATCGATGAGGACACAGCTTGAGTTCGTCAGGCGAGTTACGGAAGACACGGAT 810
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 299 GATCTCCGAGTCGATGATGATTCGCTTCAGTTCTTCCATCAGCGACTGCTGGACGAAGGC 240
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 811 AGCCTTTGGCAAGGGCCCACTGGTTTATTTCATCAGGTTGTCGACGAGCGCTGCTTGAA 870
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 239 CTCCTTCGGATCATGGTGCCTCTTCTGTAGTTGGNGTTGTGGGTACCGCAGCTTGCC 180
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 871 ACCCTACCGGAATACGAAAT 890
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 179 ATCCGCGCGGAATCGAACT 160
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

RESULT 15
ABL15160/C
ID ABL15160 standard; cDNA; 2738 BP.
XX AC ABL15160;
XX XX
XX DT 26-MAR-2002 (first entry)
XX DE Drosophila melanogaster expressed polynucleotide SEQ ID NO 39962.
XX KW Drosophila; developmental biology; cell signalling; insecticide;
XX OS pharmaceutical; gene; ss.
XX OS Drosophila melanogaster.
XX PN WO200171042-A2.
XX PD 27-SEP-2001.
XX XX
```

```
XX 23-MAR-2001; 2001WO-US009231.
XX PF 23-MAR-2000; 2000US-0191637P.
XX PR 11-JUL-2000; 2000US-00614150.
XX XX
XX PA (PEXE ) PE CORP NY.
XX PI Venter JC, Adams M, Li PWD, Myers EW;
XX DR WPI; 2001-656860/75.
XX DR P-PSDB; ABB71057.
XX PT New isolated nucleic acid detection reagent for detecting 1000 or more
XX PT genes from Drosophila and for elucidating cell signaling and cell-cell
XX PT interactions.
XX PS Claim 1; SEQ ID NO 39962; 21pp + Sequence Listing; English.
XX CC The invention relates to an isolated nucleic acid detection reagent
XX CC capable of detecting 1000 or more genes from Drosophila. The invention is
XX CC useful in developmental biology and in elucidating cell signalling and
XX CC cell-cell interactions in higher eukaryotes for the development of
XX CC insecticides, therapeutics and pharmaceutical drugs. The invention
XX CC discloses genomic DNA sequences (ABL16176-ABL30511), expressed DNA
XX CC sequences (ABL01840-ABL16175) and the encoded proteins (ABB57737-
XX CC ABB72072). The sequence data for this patent did not form part of the
XX CC printed specification, but was obtained in electronic format directly
XX CC from WIPO at ftp.wipo.int/pub/published_pct_sequences
XX SQ Sequence 2738 BP; 738 A; 643 C; 659 G; 698 T; 0 U; 0 Other;

Query Match          3.8%; Score 37.6; DB 4; Length 2738;
Best Local Similarity 54.3%; Pred. No. 0.41;
Matches 76; Conservative 0; Mismatches 64; Indels 0; Gaps 0;

QY 751 GATATCGACATCGATGAGGACACAGCTTGAGTTCGTCAGGCGAGTTACGGAAGACACGGAT 810
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 1299 GATCTCCGAGTCGATGATGATTCGCTTCAGTTCTTCCATCAGCGACTGCTGGACGAAGGC 1240
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 811 AGCCTTTGGCAAGGGCCCACTGGTTTATTTCATCAGGTTGTCGACGAGCGCTGCTTGAA 870
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 1239 CTCCTTCGGATCATGGTGCCTCTTCTGTAGTTGGNGTTGTGGGTACCGCAGCTTGCC 1180
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 871 ACCCTACCGGAATACGAAAT 890
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 1179 ATCCGCGCGGAATCGAACT 1160
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

Search completed: June 27, 2004, 05:02:43
Job time : 435.15 secs
```

GenCore version 5.1.6
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: June 27, 2004, 04:30:32 ; Search time 2839.31 Seconds
(without alignments)
10412.231 Million cell updates/sec

Title: US-10-658-691-3
Perfect score: 990
Sequence: 1 agtaccacactcaaatgga.....attttgaagcattttcttaa 990

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 27513289 seqs, 14931090276 residues

Total number of hits satisfying chosen parameters: 55026578

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : EST:*
1: em_estba:*
2: em_esthum:*
3: em_estin:*
4: em_estmu:*
5: em_estov:*
6: em_estpl:*
7: em_estro:*
8: em_hc:*
9: gb_est1:*
10: gb_est2:*
11: gb_hc:*
12: gb_est3:*
13: gb_est4:*
14: gb_est5:*
15: em_estfun:*
16: em_estom:*
17: em_gss_hum:*
18: em_gss_inv:*
19: em_gss_pla:*
20: em_gss_vrt:*
21: em_gss_fun:*
22: em_gss_lam:*
23: em_gss_mus:*
24: em_gss_pro:*
25: em_gss_rod:*
26: em_gss_phg:*
27: em_gss_vri:*
28: gb_gss1:*
29: gb_gss2:*

Result No. Score Query % Match Length DB ID Description
1 42.8 4.3 1201 13 BX381961 BX381961
C 2 40.8 4.1 600 12 B1229988 B1229988
c 3 38.2 3.9 615 12 BJ260819 BJ260819
c 4 37.8 3.8 291 28 BH569723 BH569723

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

RESULT 1
BX381961
LOCUS BX381961
DEFINITION BX381961 Homo sapiens PLACENTA COT 25-NORMALIZED Homo sapiens CDNA clone CS0DI072YF05 3-PRIME, mRNA sequence.
ACCESSION BX381961
VERSION BX381961.1 GI:30453007
KEYWORDS EST.
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 1201)
AUTHORS Li.W.B., Gruber.C., Jessee.J. and Polayes.D.
TITLE Full-length cDNA libraries and normalization
JOURNAL Unpublished (2001)
COMMENT Contact: Genoscope
Genoscope - Centre National de Sequencage
BP 191 91006 EVRY cedex - France
Email: seqref@genoscope.cns.fr, Web : www.genoscope.cns.fr
Library was constructed by Life Technologies, a division of Invitrogen. Contact : Feng Liang Email : fliang@lifetech.com URL : http://fulllength.invitrogen.com/ Invitrogen Corporation 1600 Faraday Avenue Genoscope sequence ID : CS0DI072CC03N1.
Location/Qualifiers
1. .1201
/organism="Homo sapiens"

FEATURES
source

5	37.8	3.8	440	9	AA813209	AA813209 aj32e04.s
6	37.8	3.8	552	28	CC047819	CC047819 3591.1.98
c 7	37.8	3.8	634	12	BI367122	BI367122 RE52813.5
c 8	37.6	3.8	1176	28	BZ551949	BZ551949 pacsl-60
c 9	37.6	3.8	369	12	BI167687	BI167687 RE08123.5
c 10	37.6	3.8	441	11	AY231918	AY231918 Drosophila
c 11	37.6	3.8	441	11	AY232121	AY232121 Drosophila
c 12	37.6	3.8	517	10	BE978534	BE978534 be79a04.y
c 13	37.6	3.8	541	12	BI167742	BI167742 RE08194.5
c 14	37.6	3.8	557	12	BI483125	BI483125 RE65966.5
c 15	37.6	3.8	606	12	BI167128	BI167128 RE07446.5
c 16	37.6	3.8	645	12	BI356036	BI356036 RE42061.5
c 17	37.6	3.8	661	12	BI236552	BI236552 RE22623.5
c 18	37.6	3.8	661	12	BI367637	BI367637 RE53440.5
c 19	37.6	3.8	661	12	BI371843	BI371843 RE58850.5
c 20	37.6	3.8	663	12	BI230435	BI230435 RE14116.5
c 21	37.6	3.8	663	12	BI242569	BI242569 RE39946.5
c 22	37.6	3.8	675	13	BQ705266	BQ705266 Y1A10A04
c 23	37.6	3.8	677	12	BI632106	BI632106 SD25703.5
c 24	37.6	3.8	683	12	BI228632	BI228632 RE26236.5
c 25	37.6	3.8	684	12	BI230656	BI230656 RE14375.5
c 26	37.6	3.8	686	12	BI485460	BI485460 RE68784.5
c 27	37.6	3.8	706	12	BI371383	BI371383 RE58261.5
c 28	37.6	3.8	871	13	BQ705783	BQ705783 Y1B04F09
c 29	37.2	3.7	1111	29	CNS01497	AL103861 Drosophila
c 30	36.8	3.7	884	14	CD299445	CD299445 AGENCOURT
c 31	36.8	3.7	887	12	BI834500	BI834500 603084671
c 32	36.8	3.7	932	13	BUI13845	BUI13845 603130341
c 33	36.6	3.7	688	13	BW261927	BW261927 BW261927
c 34	36.2	3.7	657	28	BH507017	BH507017 BOHQD45TR
c 35	36.2	3.7	702	28	BH936230	BH936230 Od10F02
c 36	36.2	3.7	712	28	BH513796	BH513796 BOGUX75TR
c 37	36.2	3.7	725	28	BH730081	BH730081 BOMEE23TF
c 38	36	3.6	434	9	AL817274	AL817274 AL817274
c 39	35.6	3.6	1087	9	AL575881	AL575881 AL575881
c 40	35.4	3.6	445	10	BE495071	BE495071 WHE1267.B
c 41	35.4	3.6	577	12	BM082775	BM082775 fu25906.Y
c 42	35.4	3.6	633	13	BQ263855	BQ263855 faa18h04
c 43	35.4	3.6	649	28	BZ852591	BZ852591 ZF240.219
c 44	35.4	3.6	655	12	BI891291	BI891291 ZF637-3-0
c 45	35.4	3.6	658	12	BM184423	BM184423 fv68b06.Y

National Institute of Genetics
1111 Yata, Mishima, Shizuoka 411-8540, Japan
Tel: 81-559-81-6856
Fax: 81-559-81-6855
Email: tshini@genes.nig.ac.jp.
Location/Qualifiers
1. .615
/organism="Triticum aestivum"
/mol_type="mRNA"
/cultivar="Chinese Spring"
/db_xref="taxon:4565"
/clone="whh27all"
/tissue_type="spike at heading date"
/dev_stage="Feekes' scale 10.5"
/clone_lib="Y. Ogihara unpublished cdna library, wh_h"

FEATURES
source

ORIGIN

Query Match 3.9%; Score 38.2; DB 12; Length 615;
Best Local Similarity 49.7%; Pred. No. 9.4;
Matches 97; Conservative 0; Mismatches 98; Indels 0; Gaps 0;
QY 144 CGAAGGATGTCATCAATGATGGCGGATGTCAGGACTTCTTCGCGGATCGTGA 203
DB 471 CGACGTGGAGGTAGAACAGAGCGGTCCATGCGCGGAGTAGCGCGCAGGATGGCGC 412
QY 204 GAAGGCAACCGCATCTTTCATGTCGAGTGCCTGCTCAGACCTCGGATCAAGT 263
DB 411 CGATGGCATGAGATGCCCGGAGAGGTGAGCCCTCGTGGCGCCCGGAGGC 352
QY 264 CGCAGTGCAGGCAAGTACGTCCTCCACGATTCCTCAAGATCGAAGCGGAGTTCT 323
DB 351 CGTAGGGAGAGGCGCGGCGGATCCCTTCCAGTGTGTAAGTCTGAAGTGAAGGAGG 292
QY 324 TGAGGTCCGGCGCT 338
DB 291 TCTGTGCTGCTGCT 277

RESULT 4
BH569723/c

LOCUS BH569723 291 bp DNA linear GSS 14-DEC-2001
DEFINITION BOHQF45R BOHQ Brassica oleracea genomic clone BOHQF45, genomic survey sequence.
ACCESSION BH569723
VERSION BH569723.1 GI:17821562
KEYWORDS GSS.
SOURCE Brassica oleracea
ORGANISM Brassica oleracea
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosids; eurosids II; Brassicales; Brassicaceae; Brassica.
REFERENCE 1 (bases 1 to 291)
Town, C.D., Van Aken, S., Utterback, T., Koo, H. and Fraser, C.M.
Whole genome shotgun sequencing of Brassica oleracea
Unpublished (2001)
COMMENT Contact: Chris Town
TIGR Medical Center Drive, Rockville, MD 20850, USA.
Tel: 301-838-3523
Fax: 301-838-0208
Email: cdtown@tigr.org
DNA is from a doubled haploid provided by Tom Osborn.
Seq primer: TR
Class: sheared ends.
Location/Qualifiers
1. .291
/organism="Brassica oleracea"
/mol_type="genomic DNA"
/strain="T0100DH3"
/db_xref="taxon:3712"
/clone="BOHQF45"
/clone_lib="BOHQ"
/note="vector: pHS1; Site 1: BstXI; 2-3 kb sheared

FEATURES
source

genomic DNA inserted into pHS1 using BstXI linkers"
Query Match 3.8%; Score 37.8; DB 28; Length 291;
Best Local Similarity 47.3%; Pred. No. 8.5; Indels 0; Gaps 0;
Matches 114; Conservative 0; Mismatches 127; Indels 0; Gaps 0;
QY 736 GCTGTTCCGTGGATCGATCGACATCGATGAGCAAGCTTGAGTGGTCCAGGAGTT 795
DB 262 GCTGATTAGGGGATGAACCTTGATGTGATAAGCTCTTGTCTTGAGATAAAGCCCAT 203
QY 796 ACGAAGACACGAGTAGCCTTTGCGAAGGCGCCACTGGTTTATTCATCAGGTTGTGCAC 855
DB 202 GCTGAATACTATCATGAGAAATTTGAACGTGTGTAATTTATTTGATGTGAGATAGATT 143
QY 856 GCAGCGCTGCTTGAACCCCTACCGGAATACGAAATTTATCTTCCGCTCCACCGCCTATG 915
DB 142 GACTACTTTGTTAAACCTCGTTTGAACCTGTAGTTGATTTGTGTTCCACCTATTGCT 83
QY 916 GTCACGCTACTGTCGTAATGCTCTCGGCAAGGTTGTCACCGCATCAATTCATTTT 975
DB 82 TTGATTCATGATCATCTGAACCTCAAGAAGCATTTTCATATGATCATATCATCTTTA 23
QY 976 G 976
DB 22 G 22

RESULT 5
AA813209

LOCUS AA813209 440 bp mRNA linear EST 13-FEB-1998
DEFINITION aj32e04.s1 Soares_testis_NHT Homo sapiens cdna clone 1392030 3', mRNA sequence.
ACCESSION AA813209
VERSION AA813209.1 GI:2883194
KEYWORDS EST.
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 440)
NCI-CCAP http://www.ncbi.nlm.nih.gov/ncicgap.
AUTHORS National Cancer Institute, Cancer Genome Anatomy Project (CGAP), Tumor Gene Index
TITLE Unpublished (1997)
JOURNAL Contact: Robert Strausberg, Ph.D.
COMMENT Email: cgapbs-remail.nih.gov
CDNA Library Preparation: M. Bento Soares, Ph.D., M. Fatima Bonaldo, Ph.D.
CDNA Library Arrayed by: Greg Lennon, Ph.D.
DNA Sequencing by: Washington University Genome Sequencing Center
Clone distribution: NCI-CCAP clone distribution information can be found through the I.M.A.G.E. Consortium/LLNL at: www-bio.llnl.gov/bbrp/image/image.html

Possible reversed clone: polyT not found
Seq primer: -40m3 fwd. ET from Amersham
High quality sequence stop: 378.
Location/Qualifiers
1. .440
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/clone="1392030"
/sex="male"
/lab_host="DH10B"
/clone_lib="Soares testis NHT"
/note="Vector: p7T3b-Pac (Pharmacia) with a modified polylinker; Site 1: Not I; Site 2: Eco RI; 1st strand cdna was prepared from mRNA obtained from Clontech Laboratories, Inc., and primed with a Not I - oligo (dT) primer [5]
TGTTACCAATCTGAAGTGGAGCGGCCCCCAATTTTTTTTTTTTTT 3'.

FEATURES
source

Best Local Similarity 51.5%; Pred. No. 13;
Matches 87; Conservative 0; Mismatches 82; Indels 0; Gaps 0;

QY 722 GAACTCCGCGCGAGTGTTCGGTGATCGATATCGATCGATAGGACAGCTTGAGG 781
Db 353 GAGCTCAGGCGAGATCGTCTCTCGATGATCTCCGATCGATGATGATTCCTCAT 294
QY 782 TCCTCCAGGCGAGTACGGAAGACACCGATAGCCTTTGGCAAGGCCCATGCTTTTATTC 841
Db 293 TCCTCCATCAGCGACTGCTGTCAGCAAGGCCCTCTCGGATCATGCTGCTCTGTAG 234
QY 842 ATCAGGTTTCGAGCGAGCGCTCTTCAACCTTACCGGATACGGAAT 890
Db 233 TTGGAGTGTGTGGTACCGCAGCTTGCCTCCGCGCGGAATCGAACT 185

RESULT 8
BZ551949/c 1176 bp DNA linear GSS 17-DEC-2002
LOCUS
DEFINITION
pacsi-60_3407.xl pacsi-60 Pseudomonas aeruginosa genomic clone
pacsi-60_3407, genomic survey sequence.
ACCESSION
BZ551949
VERSION
BZ551949.1 GI:27155788
KEYWORDS
GSS.
SOURCE
Pseudomonas aeruginosa
ORGANISM
Pseudomonas aeruginosa
Bacteria; Proteobacteria; Gammaproteobacteria; Pseudomonadales;
Pseudomonadaceae; Pseudomonas.
REFERENCE
1 (bases 1 to 1176)
AUTHORS
Spencer, D.H., Raymond, C.K., Smith, E.E., Sims, E.E., Hastings, M.,
Burns, J.L., Kaul, R. and Olsen, M.V.
TITLE
Whole-Genome-Sequence variation among multiple isolates of
Pseudomonas aeruginosa library
JOURNAL
J. Bacteriol. (2002) in press
COMMENT
Contact: Chris K. Raymond
Genome Center
University of Washington
Box 352145, Seattle, WA 98105-2145, USA
Tel: 2062216954
Fax: 2066857244
Email: craymond@u.washington.edu
Class: shotgun.
Location/Qualifiers
1. .1176
/organism="Pseudomonas aeruginosa"
/mol_type="genomic DNA"
/strain="1-60"
/db_xref="taxon:287"
/clone="pacsi-60_3407"
/clone_lib="pacsi-60"
/note="clinical isolate 1-60 Whole genomic shotgun
library."

ORIGIN
Query Match 3.8%; Score 37.8; DB 28; Length 1176;
Best Local Similarity 58.4%; Pred. No. 17;
Matches 66; Conservative 0; Mismatches 47; Indels 0; Gaps 0;

QY 389 TCCTCCCGCGCGAGTCTTCCTAGTAGAGGCGAGAGCATATGGAGTCTTATTAAG 501
Db 198 TCGCAACGCGCGGTCCACTGTGTCGACGAGAGCATGTCGGCGGTGGCGCGGAC 139
QY 449 ATTCAATGGCGAATTAAAGAACCCCGAGGCATATGGAGTCTTATTAAG 501
Db 138 AGTTGCGCGTTCATCTCGCAGCAATGGAAGGCTAGAAGGAATTCGATCAAG 86

RESULT 9
B1167687/c 369 bp mRNA linear EST 09-JUL-2001
LOCUS
DEFINITION
RE08123.5prime RE Drosophila melanogaster normalized Embryo pf1c-1
Drosophila melanogaster cDNA clone RE08123 5 similar to mago:
FBan0009401 located on: 2R 57C2-57C2; 04/11/2001, mRNA sequence.

ACCESSION
B1167687 GI:14633494
VERSION
EST.
KEYWORDS
SOURCE
ORGANISM
Drosophila melanogaster (fruit fly)
Drosophila melanogaster
Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
Ephydroidea; Drosophilidae; Drosophila.
1 (bases 1 to 369)
AUTHORS
Stapleton, M., Brokstein, P., Hong, L., Tyler, D., Berman, B.,
Carlson, J., Champe, M., Chavez, C., Dorsett, V., Farfan, D., Frise, E.,
George, R., Gonzalez, M., Guarin, H., Harris, N., Li, P., Liao, G.,
Mizra, S., Mungall, C.J., Nunoo, J., Pacleb, J., Paragas, V., Park, S.,
Phouanavong, S., Wan, K., Yu, C., Lewis, S.E., Celniker, S. and
Rubin, G.M.
TITLE
BDGP/HIMI RE Drosophila EST Project
JOURNAL
Unpublished (2001)
COMMENT
Contact: Stapleton, M.
BDGP
Lawrence Berkeley National Lab
One Cyclotron Rd, Berkeley, CA 94720, USA
Fax: 510 486 6798
Email: http://www.fruitfly.org/EST, est@fruitfly.berkeley.edu
hit genomic AE003453: arm:2R [16010240..16312786]
estimated-cyto:57B20-57D11: 04/11/2001
Plate: RE.81 row: B column: 11
High quality sequence stop: 315.
Location/Qualifiers
1. .369
/organism="Drosophila melanogaster"
/mol_type="mRNA"
/db_xref="taxon:7227"
/clone="RE08123"
/sex="male and female"
/dev_stage="0-24 hours mixed stage embryonic"
/lab_host="DHS-alpha Iona"
/clone_lib="RE Drosophila melanogaster normalized Embryo
pf1c-1"
/note="Organ: embryo; Vector: pFlc1; Site 1: XhoI; Site 2:
BamHI; Library was kindly generated by Piero Carninci at
the RIKEN. The library was normalized and excised using
Cre recombinase. Plasmid cDNA library."
Query Match 3.8%; Score 37.6; DB 12; Length 369;
Best Local Similarity 54.3%; Pred. No. 11;
Matches 76; Conservative 0; Mismatches 64; Indels 0; Gaps 0;

QY 751 GATATCGACATCGATGAGCAAGCTTGAGTTCGTCGAGGAGTACGGAACACGCGAT 810
Db 322 GATCTCCGAGTCGATGATTCGCTTCAGTTCTTCCATCAGGACTGTTGACGAAGCC 263
QY 811 AGCCTTTGGCAAGGCCACCTGTTTATTCATCAGGTTGTCAGCGCGCTGCTTGA 870
Db 262 CTCCTTGGGATCATGTTGCTCTTGTAGTTGAGTTGTTGGCTACCGCAGCTTGC 203
QY 871 ACCCTACCGAATACGAAT 890
Db 202 ATCCGCGCGGAACCTCGAAT 183

RESULT 10
AY231918/c 441 bp mRNA linear HTC 04-NOV-2003
LOCUS
DEFINITION
Drosophila yakuba clone yak-em_mago mRNA sequence.
ACCESSION
AY231918
VERSION
AY231918.1 GI:38048076
KEYWORDS
HTC.
SOURCE
Drosophila yakuba
ORGANISM
Drosophila yakuba
Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
Ephydroidea; Drosophilidae; Drosophila.

```

REFERENCE
1 (bases 1 to 441)
Domazet-Loso, T. and Tautz, D.
TITLE
An evolutionary analysis of orphan genes in Drosophila
JOURNAL
Genome Res. 13 (10), 2213-2219 (2003)
MEDLINE
22887302
PubMed
14525923
REFERENCE
2 (bases 1 to 441)
Domazet-Loso, T. and Tautz, D.
TITLE
Direct Submission
Submitted (07-FEB-2003) Evolution Genetics, Institute for Genetics,
Weyertal 121, Cologne 50931, Germany
Location/Qualifiers
1..441
/organism="Drosophila yakuba"
/mol_type="mRNA"
/db_xref="taxon:7245"
/cisrcs="yak-em_mago"
<1..>441
/codon_start=1
/product="similar to Drosophila melanogaster mago"
/protein_id="AAR09941.1"
/db_xref="GI:38048077"
/translation="MSTEDFLYRYVGHKGKFGHEFLFEFRPDGKLRVANNNSYKND
TMRKEAFVQSMVEELKRIIDSEIMQEDDLWPDPDRVGRQELSLIVIGDEHSITFT
SKTGSLSLDVNRKSDPEGLRCFFYVLQDLKCLVPSLIGLHPKIKPI"
FEATURES
source
CDS
ORIGIN

```

	Query Match	3.8%;	Score 37.6;	DB 11;	Length 441;
	Best Local Similarity	54.3%;	Pred. No. 12;		
	Matches 76;	Conservative 0;	Mismatches 64;	Indels 0;	Gaps 0;
Qy	751	GATATCGACATCGATGAGCACAGCTTGAGGTGCTCCAGGCAGTTACGGAGACACCGAT	810		
Db	210	GATCTCGAATCGATGATGATCCGCTTTAGCTCTCCATCAGGACCTGGTGGACGAAGCC	151		
Qy	811	AGCCTTTGGCAAGGCGCCACCTGGTTTTATTTCATCAGGTTTGTGCAAGCAGAGGCTCTCTTGAA	870		
Db	150	CTCCTTCGGATCATGTTGTGCTTCTGTAGTTGGAATTGTGGCGTACCGCAGCTTGCC	91		
Qy	871	ACCCTACCGGAATACGAAT	890		
Db	90	ATCCGGCGGAACTCGAACT	71		

RESULT 11					
AY232121/c					
LOCUS	AY232121	441 bp	mRNA	linear	
DEFINITION	Drosophila yakuba clone yak-ad_mago mRNA sequence.				
ACCESSION	AY232121				
VERSION	AY232121.1	GI:38048482			
KEYWORDS	HTC.				

SOURCE	ORGANISM
rice	<i>Drosophila yakuba</i>
	<i>Drosophila yakuba</i>
	Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha; Ephydroidea; Drosophilidae; <i>Drosophila</i> .

REFERENCE
1 (bases 1 to 441)
AUTHORS Domazet-Loso, T. and Tautz, D.
TITLE An evolutionary analysis of orphan genes in *Drosophila*
JOURNAL Genome Res. 13 (10), 2213-2219 (2003)
MEDLINE 22887302
PUBMED 14525923

1425023
REFERENCE
2 (bases 1 to 441)
Domazet-Lošo, T. and Tautz, D.
AUTHORS
Direct Submission
TITLE
Submitted (07-FEB-2003) Evol
JOURNAL
Weyertal 121, Cologne 50931,

```

FEATURES
source
1..441
Location/Qualifiers
/organism="Drosophila yakuba"
/mol_type="mRNA"
/db_xref="taxon:7245"
/clone="v4c-ad_maco"

```

```

CDS      <1..>441
          /codon_start=1
          /product="similar to Drosophila melanogaster mago"
          /protein_id="AA010144.1"
          /db_xref="GI:38048483"
          /translation="MTEDEDFYLRYYVGHKGFGHELEFEFRPDGKLRVANNNSYKND
          TMTKFAFVHQSWTEBKRIIIDSEIMQEDDLDPPDPDRVGROLEIVIGDEHISFTT
          SKTGSLLVDVNRSKDPEGLRCFYLYVQDLKCLVPSLGLHFKIKPI"

ORIGIN
Query Match      3.8%; Score 37.6; DB 11; Length 441;
Best Local Similarity 54.3%; Pred. No. 12;
Matches 76; Conservative 0; Mismatches 64; Indels 0; Gaps 0;

Qy 751 GATATCGACATCATGAGGCAAGCTTGAGTTCGTCCAGGCAGTTCAGGAAGACACCGAT 810
      |||||
Db 210 GATCTCGAATCATGATGATCCGCTTACCTCTCCATCAGGACTGGTGGACGAAGGC 151
      |||||

Qy 811 AGCGTTGGCAAGGCCCACTGCTTTTATTCATCAGGTGTGTCAAGCAGCGCTGCTTGAA 870
      |||||
Db 150 CTCCTTCGGGATCATGTGTGCTTGTGTAGTTGGAAATGTTGCGGTACCGAGCTTGCC 91
      |||||

Qy 871 ACCCTACCGGAATACGAAAT 890
      |||||
Db 90 ATSCGGCGCGAATCTCGAAT 71
      |||||

```

[illegible]

NIDDK, National Institutes of Health
6 Center Drive MSC 2715, Bldg 6, Rm B1-13, Bethesda, MD 20892 USA
Fax: (301) 496 5239

Email: oliver@helix.nih.gov.
<http://www.nidk.nih.gov/intram/people/boliver.htm>
 Tissue isolation and library construction performed at the National
 Institute of Diabetes and Digestive and Kidney Diseases, NIH (see
<http://www.nidk.nih.gov/intram/people/boliver.htm>). DNA sequencing
 and analyses performed by National Institutes of Health Intramural
 Sequencing Center (NISC; see <http://www.nisc.nih.gov/>).
 Plate: 79 row: a column: 04
 Seq primer: M13Rp1 reverse primer (ABI).

```

seq primer: m1sarl reverse primer (AB1).
Location/Qualifiers
1..517
/organism="Drosophila melanogaster"
/mol_type="mRNA"
/strain="V[*] w[67cl]/y"
/db_xref="taxon:7227"
/clone="bs79a04"
/sex="male"
/dev_stage="1-5 day adult"
/lab_host="SOLR (Stratagene)"
/clone_lib="Drosophila melanogaster"
/notes="Organ: testis; Vector: pBlue
Site 1: EcoR 1; Site 2: Xho I; Testis
day adult y[*] w[67cl]/y males raised
isolated using Trizol (Life Technol

```

the RIKEN. The library was normalized and excised using Cre recombinase. Plasmid cDNA library."

Query Match 3.8%; Score 37.6; DB 12; Length 541;
Best Local Similarity 54.3%; Pred. No. 13;
Matches 76; Conservative 0; Mismatches 64; Indels 0

RESULT 14
BI483125/c

	EST 27-AUG-2001
B483125/c	mRNA
LOCUS	587 bp
DEFINITION	R659366.5 prime RE Drosophila melanogaster normalized Emvrio plc-1
	Drosophila melanogaster cDNA clone R659366 similar to mago;
	Francois000401 GO:[polysatiation of the oocyte microtubule cytoskeleton
	(GO:0008103); oskar mRNA localization regulation (GO:0007317)
	located on; 2R 57C2-57C3 : 05/16/2001, mRNA sequence.

VERSION
KEYWORDS
SOURCE
ORGANISM

KEYWORDS: EST.
SOURCE: Drosophila melanogaster (fruit fly)
ORGANISM: Drosophila melanogaster

REFERENCE
AUTHORS

REFERENCE
AUTHORS

1. (bases 1 to 557)
Stapleton, M., Brokesein, P., Hong, L., Tyler, D., Berman, B.,
Carlson, J., Champe, M., Chavez, C., Dorsett, V., Farfan, D., Frise, E.,
George, R., Gonzalez, M., Guarin, H., Harris, N., Li, P., Liao, G.,
Misra, S., Mungall, C. J., Nunoo, J., Pacleb, J., Pargias, V., Park, S.,
Phouanavong, S., Wan, K., Yu, C., Lewis, S. E., Celniker, S. and
Rubin, G. M.

TITLE
JOURNAL

TITLE BDGP/HMI RE Drosophila EST Project
JOURNAL Unpublished (2001)

COMMENT: Contact: Stapleton, M. BDOF
Lawrence Berkeley National Lab
One Cyclotron Rd, Berkeley, CA 94720, USA
Fax: 510 486 6798
Email: http://www.fruitfly.org/EST_est@fruitfly.berkeley.edu
hit genomic AE003453: arm:2R [16010240,16312786]
estimated-cyto:57B20-57D11: 05/16/2001
Plate: RE.659 row: F column: 6
High quality sequence stop: 486.

FEATURES

```

FEATURES
source
1. 557
/organism="Drosophila melanogaster"
/mol_type="mRNA"
/db_xref="taxon:7227"
/clone="R365966"
/sex="male and female"
/dev_stage="0-24 hours mixed stage embryonic"
/lab_host="DH5-alpha Tcra"
/clone_lib="RG Drosophila melanogaster normalized Embryo
rfic-1"

```

ORIGEN

	Query Match	3.8%	Score 37.6;	DB 12;	Length 557;
	Best Local Similarity	54.3%;	Pred. No. 13;		
	Matches	76;	Conservative 0;	Mismatches 64;	Indels 0; Gaps 0;
Qy	751	GATATCGACATCGATGAGGCAAGCTTGAGGTCTCCAGGCAGTTACGGAGACACGGAT	810		
Db	320	GATCTCGAGTCGATGATGATTCGCTTCAGGTTCTCCATCAGGACCTGGTGAGCAAGGC	261		
Qy	811	AGCCTTTGGCAAGGCGCCACTGGTTTTATTTCATCAGGTTGTCGACGAGCGGCTCCTTGAA	870		
Db	260	CTCCTTCGGATCATGGTGTCGTTCTTGTAGTTGGAGTTGTTGCGCTACCGCAGCTTGCC	201		
Qy	871	ACCCCTACCGGAATACGAAAT	890		
Db	200	ATCGGCGCGGAACCTCGAACT	181		

RESULT 15	
B1167128/c	
LOCUS	B1167128
DEFINITION	606 bp mRNA linear EST 09-JUL-2001 R307446.5 Prime RE Drosophila melanogaster normalized Embryo pf1c-1 Drosophila melanogaster cDNA clone RE07446.5 similar to mago; Fban0009401 located on: 2R 57C2-57C2.; 04/11/2001, mRNA sequence.
ACCESSION	B1167128
VERSION	B1167128.1 GI:14632935
KEYWORDS	EST.
SOURCE	Drosophila melanogaster (fruit fly)
ORGANISM	Drosophila melanogaster
	Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha; Ephydroidea; Drosophilidae; Drosophila. 1 (bases 1 to 606)
REFERENCE	Stapleton,M., Brockstein,P., Hong,L., Tyler,D., Berman,B., Carlson,J., Champe,M., Chavez,C., Dorsett,V., Fafan,D., Frise,E., George,R., Gonzalez,M., Guarin,H., Harris,N., Li,P., Liao,G., Misra,S., Mungall,C.J., Nuroo,J., Pacleb,J., Paragas,V., Park,S., Phouenavong,S., Wan,K., Yu,C., Lewis,S.E., Celnikier,S. and Rubin,G.M.
TITLE	BDGP/HMI RE Drosophila EST Project
JOURNAL	Unpublished (2001)
COMMENT	Contact: Stapleton, M. BDGP

```

Lawrence Berkeley National Lab
One Cyclotron Rd, Berkeley, CA 94720, USA
Fax: 510 486 6798
Email: http://www.fruitfly.org/EST, est@fruitfly.berkeley.edu
hit Genomic AE003453: arm:2R [16010240..16312786]
estimated-cyto:57B30-57D11: 04/11/2001
Plate: R6.74 row: D column: 10
High quality sequence stop: 504.
Location/Qualifiers
1..606
/organism="Drosophila melanogaster"
/mol_type="mrna"
/db_xref="taxon:7227"
/clones="RE07446"
/sex="male and female"
/dev_stage="0-24 hours mixed stage embryonic"
/lab_host="DH5-alpha Tona"
/clone_lib="RE Drosophila melanogaster normalized Embryo
pfic-1"
/notes="Organ: embryo; Vector: pF1cl; Site: 1: XhoI; Site: 2:
BamHI; Library was kindly generated by Piero Carnici at
the RIKEN. The library was normalized and excised using
Cre recombinase. Plasmid cDNA library."

```

Query Match 3.8%; Score 37.6; DB 12; Length 606;
Best Local Similarity 54.3%; Pred. No. 14;
Matches 76; Conservative 0; Mismatches 64; Indels 0; Gaps 0;

Q: 751 GATATCGACATCGATGAGGACAACTTGAGTCTCTCAGGCAGTTACGGAAAGACACGGAT 810

```

Db      302  GATCTCCGAGTGCATGATGATTCTCGCTTCAGTTCTTCATCACGGACTGTGTGGACGAAGC 243
Qy      811  AGCCTTTGCGAAGGCCACCTGGTTTATTATCATCAGGTTGTGCACGGACGCGTCTTGAA 870
Db      242  CTCCTTTCGGATCATGGTCTGCTTCTGTAGTTGGAGTTGTTGGCGTACCGCAGCTTGC 183
Qy      871  ACCCTACCGGAATACGAAAT 890
Db      182  ATCCGGCCGGACTCGAACT 163

Search completed: June 27, 2004, 09:59:09
Job time : 2847.31 secs

```

Search completed: June 27, 2004, 09:59:09
Job time : 2847.31 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: June 27, 2004, 04:34:02 ; Search time 79.4159 Seconds
(without alignments)
6918.029 Million cell updates/sec

Title: US-10-658-691-3

Perfect score: 990

Sequence: 1 atgtaccacactcaaaattga.....attttgacgcattttcttaa 990

Scoring table: IDENTITY_NUC

Gapop 10.0 , Gapext 1.0

Searched: 682709 seqs, 277475446 residues

Total number of hits satisfying chosen parameters: 1365418

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

Issued Patents NA.*
1: /cgn2_6/ptodata/2/ina/5A_COMB.seq.*
2: /cgn2_6/ptodata/2/ina/5B_COMB.seq.*
3: /cgn2_6/ptodata/2/ina/6A_COMB.seq.*
4: /cgn2_6/ptodata/2/ina/6B_COMB.seq.*
5: /cgn2_6/ptodata/2/ina/6CTUS_COMB.seq.*
6: /cgn2_6/ptodata/2/ina/backfiles1.seq.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	64.6	6.5	4729	6	5171684-1
2	37.6	3.8	444	1	US-08-688-609-7
3	37.6	3.8	444	3	US-09-002-832-7
4	36.8	3.7	6379	1	US-08-499-215-1
5	35.8	3.6	5331	4	US-09-453-956-1
6	34.8	3.5	65042	4	US-09-784-316-3
7	34	3.4	1788	4	US-09-252-991A-863
8	34	3.4	2097	4	US-09-252-991A-885
9	33.4	3.4	4003	1	US-08-257-999-1
10	32.8	3.3	9785	1	US-08-315-387-1
11	32.6	3.3	474	4	US-09-621-976-18033
12	32.2	3.3	3972	4	US-09-489-039A-6774
13	32.2	3.3	8478	3	US-08-676-818-1
14	32.2	3.3	8478	4	US-09-407-549-1
15	32.2	3.3	4403765	3	US-09-103-840A-2
16	32.2	3.3	4411529	3	US-09-103-840A-1
17	32	3.2	505	4	US-09-621-976-15639
18	32	3.2	4583	3	US-08-810-720-9
19	32	3.2	23673	4	US-09-773-816-1
20	31.8	3.2	1521	4	US-09-252-991A-7131
21	31.8	3.2	2172	4	US-09-252-991A-1708
22	31.8	3.2	2241	4	US-09-252-991A-2016
23	31.8	3.2	2247	4	US-09-252-991A-1856
24	31.6	3.2	1194	4	US-09-252-991A-10286
25	31.6	3.2	2493	4	US-09-252-991A-10571
26	31.6	3.2	2367	4	US-09-252-991A-10675
27	31.6	3.2	3027	4	US-09-252-991A-10492

28 31.4 3.2 1044 4 US-09-252-991A-15076 Sequence 15076, A
29 31.4 3.2 1227 4 US-09-489-039A-1021 Sequence 1021, Ap
30 31.4 3.2 1431 4 US-09-489-039A-1187 Sequence 1187, Ap
31 31.4 3.2 1659 2 US-08-943-087-27 Sequence 27, Appl
32 31.4 3.2 2550 4 US-09-252-991A-14954 Sequence 14954, A
33 31.4 3.2 4403765 3 US-09-103-840A-2 Sequence 2, Appli
34 31.2 3.2 392 3 US-08-833-876-3 Sequence 3, Appli
35 31.2 3.2 392 4 US-09-489-054-3 Sequence 3, Appli
36 31.2 3.2 978 4 US-09-252-991A-10557 Sequence 10557, A
37 31.2 3.2 1110 4 US-09-252-991A-3674 Sequence 3674, Ap
38 31.2 3.2 1311 4 US-09-252-991A-10411 Sequence 10411, A
39 31.2 3.2 1491 3 US-08-833-876-1 Sequence 1, Appli
40 31.2 3.2 1491 4 US-09-483-054-1 Sequence 1, Appli
41 31.2 3.2 1566 4 US-09-252-991A-10750 Sequence 10750, A
42 31.2 3.2 1593 4 US-09-252-991A-3716 Sequence 3716, Ap
43 31.2 3.2 1800 4 US-09-252-991A-3797 Sequence 3797, Ap
44 31.2 3.2 1833 4 US-09-252-991A-3838 Sequence 3838, Ap
45 31.2 3.2 3232 4 US-08-961-527-144 Sequence 144, App

ALIGNMENTS

RESULT 1
5171684-1
; Patent No. 5171684
; APPLICANT: YEN, KWANG-MU; BLATT, LAWRENCE M.; KARL, MICHAEL R.
; TITLE OF INVENTION: BIOCONVERSIONS CATALYZED BY THE TOULUENE
; MONOOXYGENASE OF PSEUDOMONAS MENDOCINA KR-1
; NUMBER OF SEQUENCES: 41
; CURRENT APPLICATION DATA:
; FILING DATE: 28-SEP-1990
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 177,631
; FILING DATE: 05-APR-1998
; SEQ ID NO: 1:
; SSQ ID NO: 1:
; LENGTH: 4729
5171684-1

Query Match 6.5%; Score 64.6; DB 6; Length 4729;
Best Local Similarity 51.0%; Pred No. 1.8e-11;
Matches 179; Conservative 0; Mismatches 169; Indels 3; Gaps 1;

QY 68 CAGCACTTCTCTAATGGTATCGGATTTCCGTACGATGTGCATCGGAGGTTCCGAGTAT 127
Db 3668 CTGCTCTACGTGCTGAATTTGTTATTTCCATATGAGTAACTCAGGAGGTTCCGCGCAT 3727
QY 128 GCAATTCAGTTACTCGAAGGGAATGTCATCAATGTGCGCGGATGCTCCAGACTTT 187
Db 3728 GTAAGATCAGCTGCTTTGAGGAGAGGTTCTTAACTATGCGCTGATGACCGAGATTAG 3787
QY 188 CTTCCGAGATCGTGAGAAAGGCAACCGCCATCTTTGCATGCCAGTGGTGGCTCTCAG 247
Db 3788 CGCCCGTGAACCTCCGTAAG---AATCGTTTTTGGCGTCCAGTGCAAAACCATTTATCG 3844
QY 248 ACCTGCGGATCAAGTCGCAGTGCAGGCAAGTAGTCCCAACGATTTCAATCTCAAGAA 307
Db 3845 ACCTCAAAATTAAGGTCATTAAACGTCGGAGGAGCGTCTTACATCCCCCAACGTT 3904
QY 308 TGAAGCGGAAGTTGTTGAGGTCGCGGCGCTAACTCATGACCTGCTGCTCGTCCGATTAC 367
Db 3905 TCTCGACTCGAGTAGTTAGTAAGCGCTTCTCTGACGAGATGTTGAGCTGCGACTTG 3964
QY 368 GCATGATGGCGCAGCAAAATTTCTCCCGGCGAGTTCTGCCCTAGTAGAGG 418
Db 3965 AAGCGGAACAGAAAGTGTTGTTTCCACGAGGCAATATTTTATGTTGACG 4015

RESULT 2
US-08-688-609-7/c
; Sequence 7, Application US/08688609
; Patent No. 5807708

GENERAL INFORMATION:
APPLICANT: Faib, Dean A.
TITLE OF INVENTION: Conservin Compositions and Therapeutic Uses Therefor
NUMBER OF SEQUENCES: 14
CORRESPONDENCE ADDRESS:
ADDRESSES: LAHIVE & COCKFIELD
STREET: 60 State Street
CITY: Boston
STATE: Massachusetts
COUNTRY: USA
ZIP: 02109-1875

COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/002,832
FILING DATE:
PRIOR APPLICATION DATA:
APPLICATION NUMBER:

FILING DATE:
ATTORNEY/AGENT INFORMATION:
NAME: Arnold, Beth E.
REGISTRATION NUMBER: 35,430
REFERENCE/DOCKET NUMBER: MIQ-004
TELECOMMUNICATION INFORMATION:
TELEPHONE: (617)227-7400
TELEFAX: (617)227-5941
INFORMATION FOR SEQ ID NO: 7:
SEQUENCE CHARACTERISTICS:
LENGTH: 444 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: cdna
US-09-002-832-7

Query Match 3.8%; Score 37.6; DB 1; Length 444;
Best Local Similarity 54.3%; Pred. No. 0.0095;
Matches 76; Conservative 0; Mismatches 64; Indels 0; Gaps 0;

QY 751 GATATCGACATGATGAGCAAGCTTGAGTGGTCCTCCAGGCAGTTACGGAAGACACCGAT 810
DB 210 GATCTCCGAGTCGATGATTCGTTCAGTTCTTCCATCAGGACTGGTGACGAAGGC 151
QY 811 AGCCTTGGCAAGGCCCATCGTTTATTCATCAGTTGTGCGACGAGCCTGCTTGA 870
DB 150 CTCCTTGGCATGATGATTCGTTCAGTTGTGAGTTGTGGGTACCGCAGCTTGCC 91
QY 871 ACCCTACCGAATACGAAT 890
DB 90 ATCCGCGGAACTCGAACT 71

RESULT 3
US-09-002-832-7/C
Sequence 7, Application US/09002832
Patent No. 6031076
GENERAL INFORMATION:
APPLICANT: Faib, Dean A.
TITLE OF INVENTION: Conservin Compositions and Therapeutic Uses Therefor
NUMBER OF SEQUENCES: 14
CORRESPONDENCE ADDRESS:
ADDRESSES: LAHIVE & COCKFIELD
STREET: 60 State Street
CITY: Boston
STATE: Massachusetts
COUNTRY: USA
ZIP: 02109-1875

COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/688,609
FILING DATE:
PRIOR APPLICATION DATA:
APPLICATION NUMBER:

FILING DATE:
ATTORNEY/AGENT INFORMATION:
NAME: Arnold, Beth E.
REGISTRATION NUMBER: 35,430
REFERENCE/DOCKET NUMBER: MIQ-004
TELECOMMUNICATION INFORMATION:
TELEPHONE: (617)227-7400
TELEFAX: (617)227-5941
INFORMATION FOR SEQ ID NO: 7:
SEQUENCE CHARACTERISTICS:
LENGTH: 444 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: cdna
US-08-688-609-7

Query Match 3.8%; Score 37.6; DB 1; Length 444;
Best Local Similarity 54.3%; Pred. No. 0.0095;
Matches 76; Conservative 0; Mismatches 64; Indels 0; Gaps 0;

QY 751 GATATCGACATGATGAGCAAGCTTGAGTGGTCCTCCAGGCAGTTACGGAAGACACCGAT 810
DB 210 GATCTCCGAGTCGATGATTCGTTCAGTTCTTCCATCAGGACTGGTGACGAAGGC 151
QY 811 AGCCTTGGCAAGGCCCATCGTTTATTCATCAGTTGTGCGACGAGCCTGCTTGA 870
DB 150 CTCCTTGGCATGATGATTCGTTCAGTTGTGAGTTGTGGGTACCGCAGCTTGCC 91
QY 871 ACCCTACCGAATACGAAT 890
DB 90 ATCCGCGGAACTCGAACT 71

GENERAL INFORMATION:
APPLICANT: Faib, Dean A.
TITLE OF INVENTION: Conservin Compositions and Therapeutic Uses Therefor
NUMBER OF SEQUENCES: 14
CORRESPONDENCE ADDRESS:
ADDRESSES: LAHIVE & COCKFIELD
STREET: 60 State Street
CITY: Boston
STATE: Massachusetts
COUNTRY: USA
ZIP: 02109-1875

COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/002,832
FILING DATE:
PRIOR APPLICATION DATA:
APPLICATION NUMBER:

FILING DATE:
ATTORNEY/AGENT INFORMATION:
NAME: Arnold, Beth E.
REGISTRATION NUMBER: 35,430
REFERENCE/DOCKET NUMBER: MIQ-004
TELECOMMUNICATION INFORMATION:
TELEPHONE: (617)227-7400
TELEFAX: (617)227-5941
INFORMATION FOR SEQ ID NO: 7:
SEQUENCE CHARACTERISTICS:
LENGTH: 444 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: cdna
US-09-002-832-7

Query Match 3.8%; Score 37.6; DB 3; Length 444;
Best Local Similarity 54.3%; Pred. No. 0.0095;
Matches 76; Conservative 0; Mismatches 64; Indels 0; Gaps 0;

QY 751 GATATCGACATGATGAGCAAGCTTGAGTGGTCCTCCAGGCAGTTACGGAAGACACCGAT 810
DB 210 GATCTCCGAGTCGATGATTCGTTCAGTTCTTCCATCAGGACTGGTGACGAAGGC 151
QY 811 AGCCTTGGCAAGGCCCATCGTTTATTCATCAGTTGTGCGACGAGCCTGCTTGA 870
DB 150 CTCCTTGGCATGATGATTCGTTCAGTTGTGAGTTGTGGGTACCGCAGCTTGCC 91
QY 871 ACCCTACCGAATACGAAT 890
DB 90 ATCCGCGGAACTCGAACT 71

GENERAL INFORMATION:
APPLICANT: Faib, Dean A.
TITLE OF INVENTION: Conservin Compositions and Therapeutic Uses Therefor
NUMBER OF SEQUENCES: 14
CORRESPONDENCE ADDRESS:
ADDRESSES: LAHIVE & COCKFIELD
STREET: 60 State Street
CITY: Boston
STATE: Massachusetts
COUNTRY: USA
ZIP: 02109-1875

COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/499,215
FILING DATE:
CLASSIFICATION: 435
PRIOR APPLICATION DATA:

APPLICATION NUMBER: JP Hei-6-179689
FILING DATE: 08-JUL-1994
INFORMATION FOR SEQ ID NO: 1:
SEQUENCE CHARACTERISTICS:
LENGTH: 6379 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: DNA (genomic)
HYPOTHETICAL: NO
ORIGINAL SOURCE:
STRAIN: No. 5612204ardia corallina B-276 (FERM P-4094; FERM
STRAIN: BP-5124; ATCC 31338)
FEATURE:
NAME/KEY: CDS
LOCATION: 910..1935
OTHER INFORMATION: /product= "amoA"
FEATURE:
NAME/KEY: CDS
LOCATION: 1935..2285
OTHER INFORMATION: /product= "amoB"
FEATURE:
NAME/KEY: CDS
LOCATION: 2300..3802
OTHER INFORMATION: /product= "amoC"
FEATURE:
NAME/KEY: CDS
LOCATION: 3805..4830
OTHER INFORMATION: /product= "amoD"
US-08-499-215-1

Query Match 3.7%; Score 36.8; DB 1; Length 6379;
Best Local Similarity 58.0%; Pred. No. 0.1;
Matches 65; Conservative 0; Mismatches 47; Indels 0; Gaps 0;
QY 872 CCTACCGGATACGAATTATCTCGCGTCCACCGCTATGTCGACGCTACTGTCC 931
DB 4691 CCGACTGGCGCATGCTACTCTGTGGGGCGCCACCCATGATCGACGCGCTGTGC 4750
QY 932 GTATGCTGTGGCAAGGTTTCCACGCGATCAAAATTCATTTTGACGCAAT 983
DB 4751 CGCTGCTGTGGAGGGGGGTGGCGCCACGCAATCTACTACGACGCAAT 4802

RESULT 5
US-09-453-956-1
Sequence 1, Application US/09453956
Patent No. 6472191
GENERAL INFORMATION:
APPLICANT: Yano, Tetsuya; No. 6472191oto, Tsuyoshi; Imamura, Takeshi;
APPLICANT: Canon Kabushiki Kaisha
TITLE OF INVENTION: DNA Fragment Carrying Toluene Monooxygenase Gene,
TITLE OF INVENTION: Recombinant Plasmid, Transformed Microorganism,
TITLE OF INVENTION: Method for Degrading Halogenated Aliphatic Hydrocarbon
TITLE OF INVENTION: Compounds and Aromatic Compounds, and
TITLE OF INVENTION: Method for Environmental Remediation
FILE REFERENCE: CF014074US
CURRENT APPLICATION NUMBER: US/09/453,956
CURRENT FILING DATE: 1999-12-03
EARLIER APPLICATION NUMBER: JP P1998-344506
EARLIER FILING DATE: 1998-12-03
NUMBER OF SEQ ID NOS: 14
SOFTWARE: Patent In Ver. 2.1
SEQ ID NO 1
LENGTH: 5331
TYPE: DNA
ORGANISM: Ralstonia eutropha
FEATURE:
NAME/KEY: CDS
LOCATION: (200)..(445)
OTHER INFORMATION: tomK
NAME/KEY: CDS

LOCATION: (446)..(1441)
OTHER INFORMATION: tomL
FEATURE:
NAME/KEY: CDS
LOCATION: (1474)..(1743)
OTHER INFORMATION: tomM
FEATURE:
NAME/KEY: CDS
LOCATION: (1781)..(3328)
OTHER INFORMATION: tomN
FEATURE:
NAME/KEY: CDS
LOCATION: (3345)..(3701)
OTHER INFORMATION: tomO
FEATURE:
NAME/KEY: CDS
LOCATION: (3741)..(4799)
OTHER INFORMATION: tomP
FEATURE:
NAME/KEY: CDS
LOCATION: (4801)..(5136)
OTHER INFORMATION: tomQ
US-09-453-956-1
Query Match 3.6%; Score 35.8; DB 4; Length 5331;
Best Local Similarity 43.4%; Pred. No. 0.2;
Matches 279; Conservative 0; Mismatches 352; Indels 12; Gaps 2;
QY 48 AGGGAAGAGCCTGTGGTCTCAGCACCTTGCTAAATCGATCGAATTCGGTACGAGTGTC 107
DB 3794 AGGCGAGACCATTCCTGATCGCGCTCGGCCAAGGCATCTATATCCGCGCTGTG 3853
QY 108 ATCGGAGGTTGCGAGATGCAAAATTCAGTTACTCGAAGGATATGTCATCAATGTC 167
DB 3854 CCACGCGCTGTGCGGGAGCTGCAAGGTCTCGGTCTCTGACGCGAGCGCCCTGGGCGA 3913
QY 168 GCCGATGCTCCAGGACTTTCTTCGCGAGATCTGTGAGAAGGGCAACCCGCTATCTTGCATG 227
DB 3914 GGCCTACCGCTGCGGTGATGATGATTTGAGCGCGAGAGGCGCTGGCGTGTG 3973
QY 228 CCAGTGCCTGCTCTCAGACCTCGGATCAAAATTCAGTTCGAGTCAAGTCAAGTTCCTCC 287
DB 3974 CGCGACGCTGCAGGCGGATACCCACCATCGAGGCCGATGTCCACAGAGACCCCGAGCGGA 4033
QY 288 AACGATTCCTCAATCTCAAGATGGAAGCGGAAGTTGTTGAGGTTCGGGGCTAACTCATGA 347
DB 4034 GATTATCCCGGTGCGGGATTTTCGAGGCGGACGATGATGTCATCGACAGCTCACCCGAC 4093
QY 348 CCGTGTGTCGCGGATTACGCACTGATGGGCGAGCAAAATTTCTTCCCGCGGCGAGTTCTG 407
DB 4094 CATCAAGGCGATCCGCTGCGGCTCGCGAGCGGATGCGTTTCCAGGCGGGGCGAGTACGT 4153
QY 408 CCTAGTAGGCGAGCAGTTCGCGAGGCTGTTTCGCGCATATTCATATGGCGAATTTAA 467
DB 4154 CCAGTTCGAG---ATCCCGGGCTTGGGCGAGACCCGCGCTTTCTCGATCGCAACGCGCC 4210
QY 468 GAACCCCGA-----AGGCATATGGGAGTTCTATATTAAGAGGTTACCCACAGGAGC 518
DB 4211 GCGGAGCTGCGCGGACCGCGAGATCGAGCTGAACGTCGCGGAGGTTCGCGGGCGGCT 4270
QY 519 ATTAGTCTTGGCTTTTCGAAAATAAGAAAGAGCGCTGCTCTATTTTTCAGCGGACC 578
DB 4271 TGACACCGCTACCTGACGAGCTCGCCCGGGGATCGCGTGGCTTGTCCGAGCC 4330
QY 579 AATGGGCACATCTTTCTTCCTCCAGGAGACCGCGCGAAGAGTCTTTGATTTGGGCGGCG 638
DB 4331 CTATGGCCGCTTCTTGTGCGCGCTCGCGCGGCTCGCGGCTGCGGATGATCTTCATGGCGGCG 4390
QY 639 TGCGGCGCTCTGATGCGGCGCTATTTGACAGCGGCTCGATG 681
DB 4391 CTGCGGCTGTGAGCCCGCGCTCCATGATGTCGACCTGCTG 4433

RESULT 6

US-09-784-316-3
; Sequence 3, Application US/09784316
; Patent No. 6461843
; GENERAL INFORMATION:
; APPLICANT: WEI, Ming-Hui et al.
; TITLE OF INVENTION: ISOLATED HUMAN ENZYME PROTEINS, NUCLEIC
; TITLE OF INVENTION: ACID MOLECULES ENCODING HUMAN ENZYME PROTEINS, AND USES
; TITLE OF INVENTION: THREBOF
; FILE REFERENCE: CLO01139
; CURRENT APPLICATION NUMBER: US/09/784,316
; CURRENT FILING DATE: 2001-02-16
; NUMBER OF SEQ ID NOS: 5
; SOFTWARE: PastSeq for Windows Version 4.0
; SEQ ID NO 3
; LENGTH: 65042
; TYPE: DNA
; ORGANISM: Human
; NAME/KEY: misc feature
; LOCATION: (1)-(65042)
; OTHER INFORMATION: n = A,T,C or G
US-09-784-316-3

Query Match 3.5%; Score 34.8; DB 4; Length 65042;
Best Local Similarity 62.8%; Pred. No. 2.2;
Matches 54; Conservative 0; Mismatches 32; Indels 0; Gaps 0;
QY 534 TTTCGAAATAGAAAGAGCGCTCGTCTATTTTTCACGGGACCAATGGGCACATCTTT 593
DB 6894 TTTTGAATTAAGAAAGGAGAGAGTCAACAGTAGAAGACTAGACCTCTGGGCAATACAAT 6953
QY 594 CTTCCGTCACAGGACCGCGCGAAAGA 619
DB 6954 CTTACAGATTAAGCAGCAGCTGTAGA 6979

RESULT 7

US-09-252-991A-863
; Sequence 863, Application US/09252991A
; Patent No. 6551795
; GENERAL INFORMATION:
; APPLICANT: Marc J. Rubenfield et al.
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS
; TITLE OF INVENTION: AERUGINOSA FOR DIAGNOSTICS AND THERAPEUTICS
; FILE REFERENCE: 107196.136
; CURRENT APPLICATION NUMBER: US/09/252,991A
; CURRENT FILING DATE: 1999-02-18
; PRIOR APPLICATION NUMBER: US 60/074,788
; PRIOR FILING DATE: 1998-02-18
; PRIOR APPLICATION NUMBER: US 60/094,190
; PRIOR FILING DATE: 1998-07-27
; NUMBER OF SEQ ID NOS: 33142
; SEQ ID NO 863
; LENGTH: 1788
; TYPE: DNA
; ORGANISM: Pseudomonas aeruginosa
US-09-252-991A-863

Query Match 3.4%; Score 34; DB 4; Length 1788;
Best Local Similarity 56.1%; Pred. No. 0.41;
Matches 64; Conservative 0; Mismatches 50; Indels 0; Gaps 0;
QY 688 ACAGACAGCGGTAAAGTTGTTCTACGGCTCAAGAACTCCGGCGCAGCGTGTTCGGTGG 747
DB 116 ACCGTACGCCGGCAAGCAGTTCTACTGCTTCGGCTGCGCGCGCGCGCAACGCG 175
QY 748 ATCGATATCGACATCGATGAGGACAAGCTTGAGTCTCGTCAGGCGAGTTACGGAA 801
DB 176 CTCGGCTTCGTCATGGACCAACAGCAGCTGGAGTTCCCCCAGGCGGTTCGAGGAA 229

RESULT 8

US-09-252-991A-885
; Sequence 885, Application US/09252991A
; Patent No. 6551795
; GENERAL INFORMATION:
; APPLICANT: Marc J. Rubenfield et al.
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS
; TITLE OF INVENTION: AERUGINOSA FOR DIAGNOSTICS AND THERAPEUTICS
; FILE REFERENCE: 107196.136
; CURRENT APPLICATION NUMBER: US/09/252,991A
; CURRENT FILING DATE: 1999-02-18
; PRIOR APPLICATION NUMBER: US 60/074,788
; PRIOR FILING DATE: 1998-02-18
; PRIOR APPLICATION NUMBER: US 60/094,190
; PRIOR FILING DATE: 1998-07-27
; NUMBER OF SEQ ID NOS: 33142
; SEQ ID NO 885
; LENGTH: 2097
; TYPE: DNA
; ORGANISM: Pseudomonas aeruginosa
US-09-252-991A-885

Query Match 3.4%; Score 34; DB 4; Length 2097;
Best Local Similarity 56.1%; Pred. No. 0.46; Indels 0; Gaps 0;
Matches 64; Conservative 0; Mismatches 50; Indels 0; Gaps 0;
QY 688 ACAGACAGCGGTAAAGTTGTTCTACGGCTCAAGAACTCCGGCGCAGCGTGTTCGGTGG 747
DB 253 ACCGTACGCCGGCAAGCAGTTCTACTGCTTCGGCTGCGCGCGCGCAACGCG 312
QY 748 ATCGATATCGACATCGATGAGGACAAGCTTGAGTCTCGTCAGGCGAGTTACGGAA 801
DB 313 CTCGGCTTCGTCATGGACCAACAGCAGCTGGAGTTCCCCCAGGCGGTTCGAGGAA 366

RESULT 9

US-08-257-999-1/c
; Sequence 1, Application US/08257999
; Patent No. 5530195
; GENERAL INFORMATION:
; APPLICANT: Kramer, Vance C.
; APPLICANT: Curtler, Thomas C.
; TITLE OF INVENTION: No. 5530195el Bacillus Thuringiensis Gene
; TITLE OF INVENTION: Encoding a Toxin Active Against Insects
; NUMBER OF SEQUENCES: 6
; CORRESPONDENCE ADDRESS:
; ADDRESS: Ciba-Geigy Corporation
; STREET: 7 Skyline Drive
; CITY: Hawthorne
; STATE: NY
; COUNTRY: USA
; ZIP: 10532
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/257,999
; FILING DATE:
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: Elmer, James Scott
; REGISTRATION NUMBER: 36,129
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 919-541-8614
; TELEFAX: 919-541-8689
; INFORMATION FOR SEQ ID NO: 1:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 4003 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: DNA (genomic)

```
/ HYPOTHETICAL: NO
/ ANTI-SENSE: NO
/ ORIGINAL SOURCE:
/ ORGANISM: Bacillus thuringiensis
/ STRAIN: Kurstaki
/ INDIVIDUAL ISOLATE: CGB316
/ FEATURE:
/ NAME/KEY: promoter
/ LOCATION: 96..124
/ OTHER INFORMATION: /function= "putative promoter"
/ OTHER INFORMATION: region"
/ FEATURE:
/ NAME/KEY: RBS
/ LOCATION: 185..190
/ FEATURE:
/ NAME/KEY: CDS
/ LOCATION: 196..3723
/ OTHER INFORMATION: /product= "Full-length CryIE(c)"
/ OTHER INFORMATION: protein"
/ FEATURE:
/ NAME/KEY: misc feature
/ LOCATION: 1191..1590
/ OTHER INFORMATION: /note= "This region of the CryIE(c)"
/ OTHER INFORMATION: DNA sequence encodes the amino acid sequence
/ OTHER INFORMATION: designated Sub-Sequence A"
/ Patent No. 5530195
/ FEATURE:
/ NAME/KEY: misc feature
/ LOCATION: 1591..2061
/ OTHER INFORMATION: /note= "This region of the CryIE(c)"
/ OTHER INFORMATION: DNA sequence encodes the amino acid sequence
/ OTHER INFORMATION: designated Sub-Sequence B"
/ Patent No. 5530195
/ US-08-257-999-1

Query Match      3.4%; Score 33.4; DB 1; Length 4003;
Best Local Similarity 54.5%; Pred. No. 1.1;
Matches 67; Conservative 0; Mismatches 56; Indels 0; Gaps 0;

QY 862 CTGTTGAAACCCCTACCGGAATACGAAATTATCTCCCGGTCCACCGCCTATGGTCGAC 921
Db 2095 CTGTTTTTAAACCGATTGATTAGAGAGTAAACAGCGCATTCACCGCCTTTTGCTC 2036

QY 922 GCTACTGCGTATGCTGCTCGGAAGGTTTCCAGCGCATCAATTCTTTTGACGCA 981
Db 2035 TTCTAGATCATATTCCTGCTCAAGGTTACTTCTGCCGGAACAAATTCATTCGATCTA 1976

QY 982 TTT 984
Db 1975 TAT 1973

RESULT 10
US-08-319-387-1
; Sequence 1, Application US/08319387
; Patent No. 5543317
; GENERAL INFORMATION:
; APPLICANT: Shields, Malcolm S.
; APPLICANT: Francesconi, Stephen C.
; TITLE OF INVENTION: Microbial Degradation of Trichloroethylene,
; TITLE OF INVENTION: Dichloroethylenes and Aromatic Pollutants
; NUMBER OF SEQUENCES: 7
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Saliwanchik & Saliwanchik
; STREET: 2421 N.W. 41st Street, Suite A-1
; CITY: Gainesville
; STATE: FL
; COUNTRY: USA
; ZIP: 32606
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
```

```
/ SOFTWARE: PatentIn Release #1.0, Version #1.25
/ CURRENT APPLICATION DATA:
/ APPLICATION NUMBER: US/08/319,387
/ FILING DATE:
/ CLASSIFICATION: 435
/ PRIOR APPLICATION DATA:
/ APPLICATION NUMBER: US 08/167,457
/ FILING DATE: 15-DEC-1993
/ CLASSIFICATION: 435
/ PRIOR APPLICATION DATA:
/ APPLICATION NUMBER: US 07/694,718
/ FILING DATE: 02-MAY-1991
/ CLASSIFICATION: 435
/ ATTORNEY/AGENT INFORMATION:
/ NAME: Whitlock, Ted W.
/ REGISTRATION NUMBER: 36,965
/ REFERENCE/DOCKET NUMBER: UWF-2
/ TELECOMMUNICATION INFORMATION:
/ TELEPHONE: 904-375-8100
/ TELEFAX: 904-372-5800
/ INFORMATION FOR SEQ ID NO: 1:
/ SEQUENCE CHARACTERISTICS:
/ LENGTH: 9785 base pairs
/ TYPE: nucleic acid
/ STRANDEDNESS: single
/ TOPOLOGY: linear
/ MOLECULE TYPE: DNA (genomic)
/ US-08-319-387-1

Query Match      3.3%; Score 32.8; DB 1; Length 9785;
Best Local Similarity 48.0%; Pred. No. 3.2;
Matches 94; Conservative 0; Mismatches 102; Indels 0; Gaps 0;

QY 487 GAGTTCATATTAAGAGGGTACCCACAGCAGATTAGTCTTGGCTTTTCGAAAATAGA 546
Db 6680 GAGCTCAATATTCGCGGGTTCGGGAGACAGGCGGTACATCATGACAACTG 6739

QY 547 AAAGAAAGCGCTCGTCTATTTTACGGGACAAATGGGCACATCTTTCTCCGCCAGG 606
Db 6740 GCGGTGGGGAACATTCCTGACGCTGACGGGTCCCTATGCGCGGTTTTTCGCGCAGGTCA 6799

QY 607 ACCGCCGGAAGAGTCTTTGATTGGCGGCGTGGCGGCTCTCGTATGCGGCCCTATT 666
Db 6800 GCCGACACCCATCGTTTTTATGACAGTGTTCAGCCCTGTGAGCCCTCGTCCGTCGATG 6859

QY 667 GCACGCGCTCGATGC 682
Db 6860 ATTCTCGACTTGTGC 6875

RESULT 11
US-09-621-976-18033
; Sequence 18033, Application US/09621976
; Patent No. 6639063
; GENERAL INFORMATION:
; APPLICANT: Dumas Milne Edwards, J.B.
; APPLICANT: Jobert, S.
; APPLICANT: Giordano, J.Y.
; TITLE OF INVENTION: ESTs and Encoded Human Proteins.
; FILE REFERENCE: GENSET.054PR2
; CURRENT APPLICATION NUMBER: US/09/621,976
; CURRENT FILING DATE: 2000-07-21
; NUMBER OF SEQ ID NOS: 19335
; SOFTWARE: Patent.pm
; SEQ ID NO 18033
; LENGTH: 474
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: 16
; OTHER INFORMATION: n=a, g, c or t
; US-09-621-976-18033
```

```

; STREET: 225 Franklin Street
; CITY: Boston
; STATE: Massachusetts
; COUNTRY: U.S.A.
; ZIP: 02110-2804
; COMPUTER READABLE FORM:
; MEDIUM TYPE: 3.5" Diskette, 1.44 Mb
; COMPUTER: IBM PS/2 Model 502 or 55SX
; OPERATING SYSTEM: MS-DOS (Version 5.0)
; SOFTWARE: Wordperfect (Version 5.1)
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/676,818
; FILING DATE: 08-JUL-1996
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/239,430
; FILING DATE: May 6, 1994
; APPLICATION NUMBER: 08/084,709
; FILING DATE: June 25, 1993
; ATTORNEY/AGENT INFORMATION:
; NAME: Freeman, John W.
; REGISTRATION NUMBER: 29,066
; REFERENCE/DOCKET NUMBER: 04599/004001
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (617) 542-5070
; TELEFAX: (617) 542-8906
; TELEX: 200154
; INFORMATION FOR SEQ ID NO: 1:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 8478
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: linear
; US-08-676-818-1

Query Match 3.3%; Score 32.2; DB 3; Length 8478;
Best Local Similarity 46.9%; Pred. No. 4.8;
Matches 100; Conservative 0; Mismatches 113; Indels 0; Gaps 0;

QY 681 GCGCGAAACAGACAGCCGCGTAAAGTTGTTCTACGGCTCAAGAACTCCGCGCGACGCTGT 740
Ddb 2008 GCACGAGAAATGTCAGCGCTGATCTGATGGTCGCGTAAAGCAATTACAGGAGGCTATTT 2067
QY 741 TCGGTGGATCGATATCGACATCGATGAGGACAGAGCTTGAGGTCTGCAGGAGGTTACGGA 800
Ddb 2069 GCCAATGCCGTTACGTTTGCCATCGAAGACATCTATAAGGCATTCTATGATGATTATGA 2127
QY 801 AGACACGATAGCCCTTTGGCAGAGGCCACCTGGTTTTATTTCATCAGGTTGTGACGCGAGC 860
Ddb 2128 AAACTAAAACCTTTTCCATGGCCATTCCTATACAGGCAATCAGCTTGGCTGTGCGGT 2187
QY 861 GCTGCTTGAACCCCTACCGGAATCGAAATTGA 893
Ddb 2188 TCGCCTTGAATCTGGCAATTATTGTAATCTGA 2220

RESULT 14
US-09-407-549-1
; Sequence 1, Application US/09407549
; Patent No. 630377
; GENERAL INFORMATION:
; APPLICANT: Bower, Stanley Grant
; APPLICANT: Perkins, John B.
; APPLICANT: Vocum, R. Rogers
; APPLICANT: Pero, Janice G.
; TITLE OF INVENTION: BIOTIN BIOSYNTHESIS IN BACILLUS
; TITLE OF INVENTION: SUBTILIS
; NUMBER OF SEQUENCES: 25
; CORRESPONDENCE ADDRESS:
; ADDRESS: Fish & Richardson P.C.
; STREET: 225 Franklin Street
; CITY: Boston
; STATE: Massachusetts

```

COUNTRY: U.S.A.
ZIP: 02110-2804
COMPUTER READABLE FORM:
MEDIUM TYPE: 3.5" Diskette, 1.44 Mb
COMPUTER: IBM PS/2 Model 50Z or 55SX
OPERATING SYSTEM: MS-DOS (Version 5.0)
SOFTWARE: WordPerfect (Version 5.1)
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/407,549
FILING DATE:
CLASSIFICATION:
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 08/239,430
FILING DATE: May 6, 1994
APPLICATION NUMBER: 08/084,709
FILING DATE: June 25, 1993
ATTORNEY/AGENT INFORMATION:
NAME: Freeman, John W.
REGISTRATION NUMBER: 29,066
REFERENCE/POCKET NUMBER: 04599/004001
TELECOMMUNICATION INFORMATION:
TELEPHONE: (617) 542-5070
TELEFAX: (617) 542-8906
TELEX: 200154
INFORMATION FOR SEQ ID NO: 1:
SEQUENCE CHARACTERISTICS:
LENGTH: 8478
TYPE: nucleic acid
STRANDEDNESS: double
TOPOLOGY: linear
US-09-407-549-1

Query Match 3.3%; Score 32.2; DB 4; Length 8478;
Best Local Similarity 46.9%; Pred. No. 4.8;
Matches 100; Conservative 0; Mismatches 113; Indels 0; Gaps 0;
QY 691 GCGGAAACAGACGCGGTAAGTTGTTCTACGGCTCAAGAACTCCGCGGACGGCTGT 740
DB 2008 GCACGAGAATGCCAGCGCTGATCTGATGGCTGCGGTAAGGCAATTACAGGAGCTATTT 2067
QY 741 TCGGTGGATCGATATCGACATCGATGAGGCAAGCTTGAGTCTCCAGGCGATTACGGA 800
DB 2068 GCCAATGCCGTTAGCTTGGCCTGAGACATCTATAAGGCATCTATGATGATTAGA 2127
QY 801 AGACAGGATAGCTTTGGCAAGGCGCCACTGGTTTATTCATCAGGTTGTCGACGAGC 860
DB 2128 AAACCTAAACACCTTTTCCATGGCCATTCCTATACAGGCAATCAGCTTGCTGTGGGT 2187
QY 861 GCTGCTTGAACCCCTACCGGAATACGAAATTA 893
DB 2188 TCGCTTGAATAATCTGGCATTATTGGAATCTGA 2220

RESULT 15
US-09-103-840A-2
Sequence 2, Application US/09103840A
Patent No. 6294328
GENERAL INFORMATION:
APPLICANT: FLEISCHMAN, Robert D.
APPLICANT: WHITE, Owen R.
APPLICANT: PRASER, Claire M.
APPLICANT: VENTER, John C.
TITLE OF INVENTION: DNA SEQUENCES FOR STRAIN ANALYSIS IN MYCOBACTERIUM
TITLE OF INVENTION: TUBERCULOSIS
FILE REFERENCE: 24366-20007.00
CURRENT APPLICATION NUMBER: US/09/103,840A
CURRENT FILING DATE: 1998-06-24
NUMBER OF SEQ ID NOS: 2
SOFTWARE: PatentIn Ver. 2.1
SEQ ID NO 2
LENGTH: 4403765
TYPE: DNA
ORGANISM: Mycobacterium tuberculosis

FEATURE:
OTHER INFORMATION: CDC 1551
OTHER INFORMATION: "n" bases at various positions throughout the sequence
OTHER INFORMATION: represent a, t, c or g
US-09-103-840A-2
Query Match 3.3%; Score 32.2; DB 3; Length 4403765;
Best Local Similarity 47.3%; Pred. No. 1.1e+02;
Matches 97; Conservative 0; Mismatches 108; Indels 0; Gaps 0;
QY 746 GGATCGATATCGACATCGATGAGGACAAAGCTTGAGGTCTCCAGGCACTTACGGAAGACA 805
DB 1825965 GAATCGTTGTCGACGGCGATGTTTGGCGCGTGAAGTGGTCCAGCGCGCACCGAGGGGC 1826024
QY 806 CGGATAGCTTTGGCAAGGCCCCACTGGTTTATTCATCAGGTTGTCGACGAGCGCTGC 865
DB 1826025 TGGCCTCGCTGGTTCGACGCGTTTCGGTTCGGACATCTCTGTCACAGGAGCGCTGGACC 1826084
QY 866 TTGAAACCCCTACCGGAATACGAAATTTATCTTCCGCGTCCACCGCTATGTCGACGCTA 925
DB 1826085 GCGAGGCGTTGGCGCGCCCAAGCGCTTTCGAGATGACGAGTCCGCGCGTGTGCTCAACGGAA 1826144
QY 926 CTGTCCGATGCTGCTCGGCAAGGG 950
DB 1826145 TCGTGACCCGCTGGTTCGCGCGCG 1826169
Search completed: June 27, 2004, 10:02:59
Job time : 96.4159 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: June 27, 2004, 07:56:42 ; Search time 456.449 Seconds

(without alignments)
9936.011 Million cell updates/sec

Title: US-10-658-691-3

Perfect score: 990

Sequence: 1 atgtaccactcaaaattga.....atttgagcatttttctaa 990

Scoring table: IDENTITY NUC

Gapop 10.0 , Gapext 1.0

Searched: 3017426 seqs, 2290544650 residues

Total number of hits satisfying chosen parameters: 6034852

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : Published Applications NA.*

- 1: /cgn2_6/ptodata/1/pubpna/US07_PUBCOMB.seq.*
- 2: /cgn2_6/ptodata/1/pubpna/PCT_NEW_PUB.seq.*
- 3: /cgn2_6/ptodata/1/pubpna/US06_PUBCOMB.seq.*
- 4: /cgn2_6/ptodata/1/pubpna/US06_PUBCOMB.seq.*
- 5: /cgn2_6/ptodata/1/pubpna/US07_NEW_PUB.seq.*
- 6: /cgn2_6/ptodata/1/pubpna/PCTUS_PUBCOMB.seq.*
- 7: /cgn2_6/ptodata/1/pubpna/US08_NEW_PUB.seq.*
- 8: /cgn2_6/ptodata/1/pubpna/US08_PUBCOMB.seq.*
- 9: /cgn2_6/ptodata/1/pubpna/US09A_PUBCOMB.seq.*
- 10: /cgn2_6/ptodata/1/pubpna/US09B_PUBCOMB.seq.*
- 11: /cgn2_6/ptodata/1/pubpna/US09C_PUBCOMB.seq.*
- 12: /cgn2_6/ptodata/1/pubpna/US09_NEW_PUB.seq.*
- 13: /cgn2_6/ptodata/1/pubpna/US09_NEW_PUB.seq.*
- 14: /cgn2_6/ptodata/1/pubpna/US10A_PUBCOMB.seq.*
- 15: /cgn2_6/ptodata/1/pubpna/US10B_PUBCOMB.seq.*
- 16: /cgn2_6/ptodata/1/pubpna/US10C_PUBCOMB.seq.*
- 17: /cgn2_6/ptodata/1/pubpna/US10_NEW_PUB.seq.*
- 18: /cgn2_6/ptodata/1/pubpna/US60_NEW_PUB.seq.*
- 19: /cgn2_6/ptodata/1/pubpna/US60_PUBCOMB.seq.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	990	100.0	990	17	US-10-658-691-3
2	54.6	5.5	12808	10	US-09-843-250-10
3	53	5.4	9706	10	US-09-843-250-5
4	49.8	5.0	4355	10	US-09-843-250-7
5	45.6	4.6	14462	10	US-09-843-250-9
6	37	3.7	1029	13	US-10-282-122A-13049
7	37	3.7	3673778	15	US-10-312-841-1
8	36.4	3.7	2394	13	US-10-282-122A-33544
9	35.8	3.6	5331	15	US-10-223-371B-1
10	35.6	3.6	632	9	US-09-864-761-20742
11	35.6	3.6	945	13	US-10-282-122A-12871
12	35.6	3.6	3981	9	US-09-864-761-3981
13	35.4	3.6	1240	16	US-10-260-238-3074
14	34.8	3.5	65042	15	US-10-229-124-3

15	34.6	3.5	1038	15	US-10-213-878-11	Sequence 11, Appl
16	34.6	3.5	1038	15	US-10-214-059-11	Sequence 11, Appl
17	34.6	3.5	1041	15	US-10-213-878-21	Sequence 21, Appl
18	34.6	3.5	1041	15	US-10-214-059-21	Sequence 21, Appl
19	34.6	3.5	12591	15	US-10-213-878-6	Sequence 6, Appl
20	34.6	3.5	12591	15	US-10-214-059-6	Sequence 6, Appl
C 21	34	3.4	888	13	US-10-398-916-25	Sequence 25, Appl
C 22	34	3.4	900	13	US-10-398-916-12	Sequence 12, Appl
C 23	34	3.4	1995	13	US-10-282-122A-30013	Sequence 30013, A
C 24	33.8	3.4	586	13	US-10-027-632-129703	Sequence 129703,
C 25	33.8	3.4	586	16	US-10-027-632-129703	Sequence 129703,
C 26	33.8	3.4	1185	16	US-10-369-493-41550	Sequence 41550, A
C 27	33.8	3.4	1711	16	US-10-369-493-41550	Sequence 581, App
C 28	33.2	3.4	1023	13	US-10-389-647-172	Sequence 172, App
C 29	33	3.3	750	15	US-10-184-644-104	Sequence 104, App
C 30	33	3.3	750	15	US-10-184-634-104	Sequence 104, App
C 31	33	3.3	912	13	US-10-282-122A-15013	Sequence 15013, A
C 32	33	3.3	2117	13	US-10-425-114-13860	Sequence 13860, A
C 33	32.8	3.3	880	16	US-10-369-493-32274	Sequence 32274, A
C 34	32.6	3.3	696	13	US-10-142-426-354	Sequence 354, App
C 35	32.6	3.3	696	15	US-10-123-155-354	Sequence 354, App
C 36	32.6	3.3	696	15	US-10-146-731-354	Sequence 354, App
C 37	32.6	3.3	696	15	US-10-140-472-354	Sequence 354, App
C 38	32.6	3.3	696	15	US-10-141-761-354	Sequence 354, App
C 39	32.6	3.3	696	15	US-10-142-885-354	Sequence 354, App
C 40	32.6	3.3	696	15	US-10-158-790-354	Sequence 354, App
C 41	32.6	3.3	696	16	US-10-137-871-354	Sequence 354, App
C 42	32.6	3.3	696	16	US-10-140-923-354	Sequence 354, App
C 43	32.6	3.3	696	16	US-10-141-756-354	Sequence 354, App
C 44	32.6	3.3	696	16	US-10-141-759-354	Sequence 354, App
C 45	32.6	3.3	696	16	US-10-140-805-354	Sequence 354, App

ALIGNMENTS

RESULT 1

US-10-658-691-3
; Sequence 3, Application US/10658691
; Publication No. US20040110258A1
; GENERAL INFORMATION:
; APPLICANT: Kayser, Kevin J.
; APPLICANT: Kilbane, John J.
; TITLE OF INVENTION: Method for Metabolizing Carbazole in Petroleum
; FILE REFERENCE: GTI-1512
; CURRENT APPLICATION NUMBER: US/10/658,691
; CURRENT FILING DATE: 2003-09-09
; PRIOR APPLICATION NUMBER: US 60/409,562
; PRIOR FILING DATE: 2002-09-10
; NUMBER OF SEQ ID NOS: 4
; SOFTWARE: PatentIn version 3.2
; SEQ ID NO 3
; LENGTH: 990
; TYPE: DNA
; ORGANISM: Pseudomonas resinovorans
US-10-658-691-3

Query Match 100.0%; Score 990; DB 17; Length 990;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 990; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy	1	ATGTACCACTCAAAATTAAGGGCAGGACCTGGGCTCAGGAGAGCCCTG	60
Db	1	ATGTACCACTCAAAATTAAGGGCAGGACCTGGGCTCAGGAGAGCCCTG	60
Qy	61	TTGGTCTCAGCACTTCTCTAATGATCGGATTTCCGTACAGTGTGCATCGGAGGTTGC	120
Db	61	TTGGTCTCAGCACTTCTCTAATGATCGGATTTCCGTACAGTGTGCATCGGAGGTTGC	120
Qy	121	GGAGTATGCAAAATTCAGTACTTCGAGGGAATGTCATGTCGCGGATGCTCCA	180
Db	121	GGAGTATGCAAAATTCAGTACTTCGAGGGAATGTCATGTCGCGGATGCTCCA	180

US-09-843-250-5

Query Match 5.4%; Score 53; DB 10; Length 9706;
Best Local Similarity 50.1%; Pred. No. 6.4e-07;
Matches 190; Conservative 0; Mismatches 180; Indels 9; Gaps 2;

QY 303 AAGAATGGAAGCGAAGTTGTTGAGGTCCGGGCGCTAACTCATGACCTCTGTCCGTGCG 362
DB 1193 AATCATCAAGGGCACAGTGTGCGAGTCGAGTCGCCCACTCAGCATATCCGTGCTTACG 1252
QY 363 ATTACGCACTGATGGCCACCAATTTCTCCCGGCCAGTTCTGCTAGTAGAGCGACA 422
DB 1253 CGTACGCTCTCAAGCCCTTCGAGTTCTCACCGGACAGTACGCACT-----GCA 1306
QY 423 GCAGTTGCCAGGGGTGTTCCGCATATTCATATGGCGAAATTTAAAGAACCCCGAAGGCAT 482
DB 1307 GTTCAGCCCTGAGCATGCGCGTCCGTATTTCATATGGCAGGTTTCCAGATGACCAAGAAT 1366
QY 483 ATGGGAGTTCTATATTAAGAGGTACCCACAGAACATTTAGTCTTGGCTTTTCGAAA 542
DB 1367 ---GGAGTTCCACATACGCAAGGTGCGGGTGGCGCGTACGAGTATGTTTTCGAACA 1423
QY 543 TAGAAAGAGGGCGTCTGTTATTTTTCAGCGGACCAATGGGCACATCTTTCTTCGTGCC 602
DB 1424 CGTCCGGAAGGTACAAGCATCAAGTTGACGGGCTCTTGTACGGCTTATCTAGCTCA 1483
QY 603 AGGACCGGCGGAAGAGTCTTTGCATTTGGCGGGTGGCGGCTCTGTATGCGGCCGC 662
DB 1484 GAAGCACACCGGACCGATGCTGTGTAGTGGCGGACCGGACTCGCACCGTGTCTGTC 1543
QY 663 TATTGACGCGCTCGCATG 681
DB 1544 GATTGTCGGCGCGCTG 1562

RESULT 4

US-09-843-250-7
; Sequence 7, Application US/09843250
; Publication No. US20030022335A1
; GENERAL INFORMATION:
; APPLICANT: Parales, R.
; APPLICANT: Gibson, D.
; APPLICANT: Resnick, S.
; APPLICANT: Lee, K.
; TITLE OF INVENTION: No. US20030022335A1 naphthalene dioxygenase and methods for the
; FILE REFERENCE: 875.006US2
; CURRENT APPLICATION NUMBER: US/09/843,250
; PRIOR FILING DATE: 2001-04-26
; PRIOR FILING DATE: 1999-10-26
; PRIOR FILING DATE: 1998-10-26
; NUMBER OF SEQ ID NOS: 65
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 7
; LENGTH: 4355
; TYPE: DNA
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: A modified DNA molecule encoding valine at the
; OTHER INFORMATION: position corresponding to the F352 amino acid in
; OTHER INFORMATION: NDO.

Query Match 5.0%; Score 49.8; DB 10; Length 4355;
Best Local Similarity 50.1%; Pred. No. 5.8e-06;
Matches 182; Conservative 0; Mismatches 172; Indels 9; Gaps 2;
QY 319 GTTGTGAGGTCCGGGCGCTAACTCATGACCTGTCTCCGTGCGAATTAACGCACTGATGG 378
DB 1159 GTGTGCGCGGTGAGTGTGCGCCCACTCAGATATCGTGTGCTACGCTACGCTGCTGAAG 1218
QY 379 CCAGCAATTTCTCCCGCGCCGCTTCTGCTAGTAGAGGACAGACAGTTGCCAGGCGTG 438

DB 1219 CCCTTCGAGTTCTCACCCGGACAGTACGCGACATT-----GCAGTTCACTCTGAGCAT 1272
QY 439 GTTCCGCAATATCAATGGCGAAATTTAAAGAACCCCGAAGCATATGGAGTTCTATATT 498
DB 1273 GCGGTCTCGTATTAATGGCAGGTCTGCCAGATACCAAGAAAT---GGAGTTCCACATA 1329
QY 499 AAGAGGTATCCACACAGGACGATTTAGTCTTGGCTTTTCGAAAATAGAAAAGAGGCGCT 558
DB 1330 CGCAAGGTGCGGGTGGCGGTAACGGAGTATGTTTTCGAGCAGCTCCGCGAAGGTACA 1389
QY 559 CGTCTATTTTTCAGCGGACCAATGGGCACATCTTTCTTCGTCAGGACCGCGCGAAG 618
DB 1390 AGCATCAAGTTGAGCGGCCACITTTGTACCGCTTATTTGCGTCAAGAACACACCGGGCG 1449
QY 619 AGTCTTTTCATTTGGCGGGTGGCGGCTCTCGTATGCGGCGCTATTTGACCGGCTCG 678
DB 1450 ATGCTCTGTGTGGCGGTGGGACCGGACTAGCACCGGTGCTGCTGATTTGTCGGGCGG 1509
QY 679 ATG 681
DB 1510 CTG 1512

RESULT 5

US-09-843-250-9
; Sequence 9, Application US/09843250
; Publication No. US20030022335A1
; GENERAL INFORMATION:
; APPLICANT: Parales, R.
; APPLICANT: Gibson, D.
; APPLICANT: Resnick, S.
; APPLICANT: Lee, K.
; TITLE OF INVENTION: No. US20030022335A1 naphthalene dioxygenase and methods for the
; FILE REFERENCE: 875.006US2
; CURRENT APPLICATION NUMBER: US/09/843,250
; PRIOR FILING DATE: 2001-04-26
; PRIOR FILING DATE: 1999-10-26
; PRIOR FILING DATE: 1998-10-26
; NUMBER OF SEQ ID NOS: 65
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 9
; LENGTH: 14462
; TYPE: DNA
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: A modified DNA molecule encoding valine at the
; OTHER INFORMATION: position corresponding to the F352 amino acid in
; OTHER INFORMATION: NDO.

Query Match 4.6%; Score 45.6; DB 10; Length 14462;
Best Local Similarity 52.0%; Pred. No. 0.00034;
Matches 102; Conservative 0; Mismatches 94; Indels 0; Gaps 0;
QY 486 GGAGTTCTATATTAAGAGGTATCCACAGGACGATTTAGTCTTGGCTTTTCGAAAATAG 545
DB 4235 GGAGTTCCACATACCAAGGTGCGGGTGGCGGCTAACTGAGTATGTTTTCGACAGCT 4294
QY 546 AAAAGAGCGCTCTGCTATTTTTCAGCGGACCAATGGGCACATCTTTCTCCGTCAGG 605
DB 4295 CCGCAAGGTACAAGCATCAAAATTGACGGGCGCACTTGGTACGGCTTATTTACGTCAGAA 4354
QY 606 GACCGCGCCGAAAGAGTCTTTTTCATTTGGCGGGCGGTCCCGGCTCTGATGCGCGCTAT 665
DB 4355 CCACACCGCGCGATGCTCTGTGTGGCGGTGGACCGGACTAGCACCGGTCTGTCGAT 4414
QY 666 TCACGCGCTCGATG 681
DB 4415 TATTCGCGGCGCTG 4430

RESULT 6
US-10-282-122A-13049
; Sequence 13049, Application US/10282122A
; Publication No. US20040029129A1
; GENERAL INFORMATION:
; APPLICANT: Wang, Liangsu
; APPLICANT: Zamudio, Carlos
; APPLICANT: Malone, Cheryl
; APPLICANT: Haselbeck, Robert
; APPLICANT: Ohlsen, Kari
; APPLICANT: Zyskind, Judith
; APPLICANT: Wall, Daniel
; APPLICANT: Trawick, John
; APPLICANT: Carr, Grant
; APPLICANT: Yamamoto, Robert
; APPLICANT: Forsyth, R.
; APPLICANT: Xu, H.
; TITLE OF INVENTION: Identification of Essential Genes in Microorganisms
; FILE REFERENCE: ELITRA.034A
; CURRENT APPLICATION NUMBER: US/10/282,122A
; CURRENT FILING DATE: 2003-02-20
; PRIOR APPLICATION NUMBER: 60/191,078
; PRIOR FILING DATE: 2000-03-21
; PRIOR APPLICATION NUMBER: 60/206,848
; PRIOR FILING DATE: 2000-05-23
; PRIOR APPLICATION NUMBER: 60/207,727
; PRIOR FILING DATE: 2000-05-26
; PRIOR APPLICATION NUMBER: 60/230,335
; PRIOR FILING DATE: 2000-09-06
; PRIOR APPLICATION NUMBER: 60/230,347
; PRIOR FILING DATE: 2000-09-09
; PRIOR APPLICATION NUMBER: 60/242,578
; PRIOR FILING DATE: 2000-10-23
; PRIOR APPLICATION NUMBER: 60/253,625
; PRIOR FILING DATE: 2000-11-27
; PRIOR APPLICATION NUMBER: 60/257,931
; PRIOR FILING DATE: 2000-12-22
; PRIOR APPLICATION NUMBER: 60/267,636
; PRIOR FILING DATE: 2001-02-09
; PRIOR APPLICATION NUMBER: 60/269,308
; PRIOR FILING DATE: 2001-02-16
; Remaining Prior Application data removed - See File Wrapper or PALM.
; NUMBER OF SEQ ID NOS: 78614
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 13049
; LENGTH: 1029
; TYPE: DNA
; ORGANISM: Burkholderia fungorum
US-10-282-122A-13049

Query Match 3.7%; Score 37; DB 13; Length 1029;
Best Local Similarity 47.8%; Pred. No. 0.097;
Matches 140; Conservative 0; Mismatches 150; Indels 3; Gaps 1;
QY 433 GCGGTGTTCCGCGATATCAATGGCGAATTTAAAGAACCCGAGGCATATGGGAGTTC 492
Db 433 GCGAAGCCGCGAGCTATTCATGCGGATGCCGCCACACGAGGCCGATCGAATG 492
QY 493 TATATTAAGAGGTACCCAGACGATTAGTCCTTGCGCTTTTCGAAATAGAAAGAA 552
Db 493 CATATCCGTACATGCGCGGTGTGCGCTTTACCGATCATGTGTCAACACCATGAAAGAG 552
QY 553 GCGCTCGTCTATTTTTCGCGGACCAATGGGCACATCTTCTTCGTCGAGGACCGGC 612
Db 553 CGGACATCTGCGCTTCGAAGCCGCTCGGCACGCTTCTTCGCGCAAGATTCGGAC 612
QY 613 CGAAGAGCTTTTCATTTGCGCGGTGCGCGGCTCT---CGTATCGGCGCGTATTCGA 669
Db 613 AAGCCGATCGTGTCTCGGTGCGGTACGGGCTTCGGCGCTTGAAGCGATCATCGAG 672
QY 670 CGGCTCGATGCGCGAAGCAGACCGGTAAGTTGTTCTACGGCTCAAG 722

Db 673 CACGCGGTGTTCAAGAACCTGAACGGCGGATGACGCTGTACTGGGGCGGACG 725
RESULT 7
US-10-312-841-1
; Sequence 1, Application US/10312841
; Publication No. US20030186277A1
; GENERAL INFORMATION:
; APPLICANT: Epigenomics AG
; TITLE OF INVENTION: Diagnose von bedeutenden genetischen Parametern innerhalb des MH
; FILE REFERENCE: E01/1208/WO
; CURRENT APPLICATION NUMBER: US/10/312,841
; CURRENT FILING DATE: 2002-12-30
; NUMBER OF SEQ ID NOS: 2
; SEQ ID NO 1
; LENGTH: 3673778
; TYPE: DNA
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: chemically treated genomic DNA (Homo sapiens)
; FEATURE:
; NAME/KEY: unsure
; LOCATION: (3294164)
US-10-312-841-1

Query Match 3.7%; Score 37; DB 15; Length 3673778;
Best Local Similarity 57.3%; Pred. No. 7.1;
Matches 67; Conservative 0; Mismatches 50; Indels 0; Gaps 0;
QY 738 TGTTCCGTGATCGAATATCGATATCGATGAGGACACGCTGAGTCGTCAGGAGTTAC 797
Db 878233 TGTTAGTTAAATCGAGGTGATTTTGAAGATTATAAGCGGAGGTGTTGGAGGTTTGGG 878292
QY 798 GGAAGACACGATAGCCCTTTGGCAAGGCCCCACTGGTTTTATTTCATCAGGTTGTGCA 854
Db 878293 TGAAGTTATGTTAGTCGTTTCGAGAGTGAAGTTTTTTTTTTAGAGTTTTTTCGA 878349

RESULT 8
US-10-282-122A-33544
; Sequence 33544, Application US/10282122A
; Publication No. US20040029129A1
; GENERAL INFORMATION:
; APPLICANT: Wang, Liangsu
; APPLICANT: Zamudio, Carlos
; APPLICANT: Malone, Cheryl
; APPLICANT: Haselbeck, Robert
; APPLICANT: Ohlsen, Kari
; APPLICANT: Zyskind, Judith
; APPLICANT: Wall, Daniel
; APPLICANT: Trawick, John
; APPLICANT: Carr, Grant
; APPLICANT: Yamamoto, Robert
; APPLICANT: Forsyth, R.
; APPLICANT: Xu, H.
; TITLE OF INVENTION: Identification of Essential Genes in Microorganisms
; FILE REFERENCE: ELITRA.034A
; CURRENT APPLICATION NUMBER: US/10/282,122A
; CURRENT FILING DATE: 2003-02-20
; PRIOR APPLICATION NUMBER: 60/191,078
; PRIOR FILING DATE: 2000-03-21
; PRIOR APPLICATION NUMBER: 60/206,848
; PRIOR FILING DATE: 2000-05-23
; PRIOR APPLICATION NUMBER: 60/207,727
; PRIOR FILING DATE: 2000-05-26
; PRIOR APPLICATION NUMBER: 60/230,335
; PRIOR FILING DATE: 2000-09-06
; PRIOR APPLICATION NUMBER: 60/230,347
; PRIOR FILING DATE: 2000-09-09
; PRIOR APPLICATION NUMBER: 60/242,578
; PRIOR FILING DATE: 2000-10-23
; PRIOR APPLICATION NUMBER: 60/253,625
; PRIOR FILING DATE: 2000-11-27

;; TITLE OF INVENTION: GENE EXPRESSION ANALYSIS BY MICROARRAY
;; FILE REFERENCE: Aeomica-X-1
;; CURRENT APPLICATION NUMBER: US/09/864,761
;; CURRENT FILING DATE: 2001-05-23
;; PRIOR APPLICATION NUMBER: US 60/180,312
;; PRIOR FILING DATE: 2000-02-04
;; PRIOR APPLICATION NUMBER: US 60/207,456
;; PRIOR FILING DATE: 2000-05-26
;; PRIOR APPLICATION NUMBER: US 09/632,366
;; PRIOR FILING DATE: 2000-08-03
;; PRIOR APPLICATION NUMBER: GB 24263.6
;; PRIOR FILING DATE: 2000-10-04
;; PRIOR APPLICATION NUMBER: US 60/236,359
;; PRIOR FILING DATE: 2000-09-27
;; PRIOR APPLICATION NUMBER: PCT/US01/00666
;; PRIOR FILING DATE: 2001-01-30
;; PRIOR APPLICATION NUMBER: PCT/US01/00667
;; PRIOR FILING DATE: 2001-01-30
;; PRIOR APPLICATION NUMBER: PCT/US01/00664
;; PRIOR FILING DATE: 2001-01-30
;; PRIOR APPLICATION NUMBER: PCT/US01/00669
;; PRIOR FILING DATE: 2001-01-30
;; PRIOR APPLICATION NUMBER: PCT/US01/00665
;; PRIOR FILING DATE: 2001-01-30
;; PRIOR APPLICATION NUMBER: PCT/US01/00668
;; PRIOR FILING DATE: 2001-01-30
;; PRIOR APPLICATION NUMBER: PCT/US01/00663
;; PRIOR FILING DATE: 2001-01-30
;; PRIOR APPLICATION NUMBER: PCT/US01/00662
;; PRIOR FILING DATE: 2001-01-30
;; PRIOR APPLICATION NUMBER: PCT/US01/00661
;; PRIOR FILING DATE: 2001-01-30
;; PRIOR APPLICATION NUMBER: PCT/US01/00670
;; PRIOR FILING DATE: 2001-01-30
;; PRIOR APPLICATION NUMBER: US 60/234,687
;; PRIOR FILING DATE: 2000-09-21
;; PRIOR APPLICATION NUMBER: US 09/608,408
;; PRIOR FILING DATE: 2000-06-30
;; PRIOR APPLICATION NUMBER: US 09/774,203
;; PRIOR FILING DATE: 2001-01-29
;; NUMBER OF SEQ ID NOS: 49117
;; SOFTWARE: Annonax Sequence Listing Engine vers. 1.1
;; SEQ ID NO 20742
;; LENGTH: 632
;; TYPE: DNA
;; ORGANISM: Homo sapiens
;; FEATURE:
;; OTHER INFORMATION: MAP TO AE000658.1
;; OTHER INFORMATION: EXPRESSED IN FETAL LIVER, SIGNAL = 8.5
;; OTHER INFORMATION: EXPRESSED IN BT474, SIGNAL = 7.1
;; OTHER INFORMATION: EXPRESSED IN ADULT LIVER, SIGNAL = 8.3
;; OTHER INFORMATION: EXPRESSED IN BRAIN, SIGNAL = 6.3
;; OTHER INFORMATION: EXPRESSED IN BONE MARROW, SIGNAL = 7.8
;; OTHER INFORMATION: EXPRESSED IN LUNG, SIGNAL = 11
;; OTHER INFORMATION: EXPRESSED IN PLACENTA, SIGNAL = 8.3
;; OTHER INFORMATION: EXPRESSED IN HELA, SIGNAL = 11
;; OTHER INFORMATION: EXPRESSED IN HEART, SIGNAL = 5.6
;; OTHER INFORMATION: EXPRESSED IN HEL100, SIGNAL = 15
;; OTHER INFORMATION: NT HIT: G5032026, EVALUE 0.00e+00
;; OTHER INFORMATION: SWISSPROT HIT: Q09028, EVALUE 1.00e-113
;; OTHER INFORMATION: EST_HUMAN HIT: BE537587.1, EVALUE 0.00e+00
US-09-864-761-20742

Query Match 3.6%; Score 35.6; DB 9; Length 632;
Best Local Similarity 52.7%; Pred. No. 0.24; Indels 0; Gaps 0;
Matches 77; Conservative 0; Mismatches 69; Indels 0; Gaps 0;
QY 692 ACAAGCGGTAAGTTGTTTACGGCTCAAGAACTCCGCGGACGCTGTTCGGTGGATCG 751
Db 254 AGATTCCAGAAAACATCTTCTACTGCTGTATGCCCGTAAGATGCTCCTTGCATCC 195
QY 752 ATATCGACATCGATAGGCAAGCTTGAGTCTGTCAGGACGATTACGGAGACCGGTA 811

Db 194 ACTTTTCCCTCTTTGGAATGGCACTGATGTCCTCCACAGGACAGATGTGTGTCATCTGAA 135
QY 812 GCCTTTGGCAAGGCCCCACACTGGTTTT 837
Db 134 GCCTAAGTAATGCCACCTGAGATT 109
RESULT 11
US-10-282-122A-12871
;; Sequence 12871, Application US/10282122A
;; Publication No. US20040029129A1
;; GENERAL INFORMATION:
;; APPLICANT: Wang, Liangsu
;; APPLICANT: Zamudio, Carlos
;; APPLICANT: Malone, Cheryl
;; APPLICANT: Haselbeck, Robert
;; APPLICANT: Ohlsen, Kari
;; APPLICANT: Zyskind, Judith
;; APPLICANT: Wall, Daniel
;; APPLICANT: Trawick, John
;; APPLICANT: Carr, Grant
;; APPLICANT: Yamamoto, Robert
;; APPLICANT: Forsyth, R.
;; APPLICANT: Xu, H.
;; TITLE OF INVENTION: Identification of Essential Genes in Microorganisms
;; FILE REFERENCE: ELITRA.034A
;; CURRENT APPLICATION NUMBER: US/10/282,122A
;; CURRENT FILING DATE: 2003-02-20
;; PRIOR APPLICATION NUMBER: 60/191,078
;; PRIOR FILING DATE: 2000-03-21
;; PRIOR APPLICATION NUMBER: 60/206,848
;; PRIOR FILING DATE: 2000-05-23
;; PRIOR APPLICATION NUMBER: 60/207,727
;; PRIOR FILING DATE: 2000-05-26
;; PRIOR APPLICATION NUMBER: 60/230,335
;; PRIOR FILING DATE: 2000-09-06
;; PRIOR APPLICATION NUMBER: 60/230,347
;; PRIOR FILING DATE: 2000-09-09
;; PRIOR APPLICATION NUMBER: 60/242,578
;; PRIOR FILING DATE: 2000-10-23
;; PRIOR APPLICATION NUMBER: 60/253,625
;; PRIOR FILING DATE: 2000-11-27
;; PRIOR APPLICATION NUMBER: 60/257,931
;; PRIOR FILING DATE: 2000-12-22
;; PRIOR APPLICATION NUMBER: 60/267,636
;; PRIOR FILING DATE: 2001-02-09
;; PRIOR APPLICATION NUMBER: 60/269,308
;; PRIOR FILING DATE: 2001-02-16
;; Remaining Prior Application data removed - See File Wrapper or PALM.
;; NUMBER OF SEQ ID NOS: 78614
;; SOFTWARE: PatentIn version 3.1
;; SEQ ID NO 12871
;; TYPE: DNA
;; LENGTH: 945
;; ORGANISM: Burkholderia fungorum
US-10-282-122A-12871

Query Match 3.6%; Score 35.6; DB 13; Length 945;
Best Local Similarity 50.6%; Pred. No. 0.29; Indels 0; Gaps 0;
Matches 86; Conservative 0; Mismatches 84; Indels 0; Gaps 0;
QY 131 AATTCGAGTTACTCGAAGGGAATGTCCAATGTGCCCGGATGCTCCAGGACTTTCTT 190
Db 302 AACTCGCGCTCTCGAACGACGCGTGTCTCTCCAGTTTCGGGAAGCGGGCGCTGGTA 361
QY 191 CGCAGATCGTGAGAGAGGCAACCGCCATCTTGCATGCCAGTGGTTCGCTCTCAGACC 250
Db 362 CCGCGGAATTTCTGCGTGGCCCGGCAATTTTCGGAAAGCGGGTGGCCCGGAAGTGC 421
QY 251 TGCGGATCAAAAGTCGCGAGTGCAGGACAAAGTACGTCCCAACGATTCCCAATC 300
Db 422 TGGCGAGGCGCTGCTCGGTGCGGAGGAGGCCATTACGGACCTGATC 471

RESULT 12

US-09-864-761-3981/c

; Sequence 3981, Application US/09864761

; Patent No. US20020048763A1

; GENERAL INFORMATION:

; APPLICANT: Penn, Sharon G.

; APPLICANT: Rank, David R.

; APPLICANT: Hanzel, David K.

; APPLICANT: Chen, Wensheng

; TITLE OF INVENTION: HUMAN GENOME-DERIVED SINGLE EXON NUCLEIC ACID PROBES USEFUL FOR

; FILE REFERENCE: Gene Expression Analysis by Microarray

; CURRENT APPLICATION NUMBER: US/09/864,761

; CURRENT FILING DATE: 2001-05-23

; PRIOR APPLICATION NUMBER: US 60/180,312

; PRIOR FILING DATE: 2000-02-04

; PRIOR APPLICATION NUMBER: US 60/207,456

; PRIOR FILING DATE: 2000-05-26

; PRIOR APPLICATION NUMBER: US 09/632,366

; PRIOR FILING DATE: 2000-08-03

; PRIOR APPLICATION NUMBER: GB 24263.6

; PRIOR FILING DATE: 2000-10-04

; PRIOR APPLICATION NUMBER: US 60/236,359

; PRIOR FILING DATE: 2000-09-27

; PRIOR APPLICATION NUMBER: PCT/US01/00666

; PRIOR FILING DATE: 2001-01-30

; PRIOR APPLICATION NUMBER: PCT/US01/00667

; PRIOR FILING DATE: 2001-01-30

; PRIOR APPLICATION NUMBER: PCT/US01/00664

; PRIOR FILING DATE: 2001-01-30

; PRIOR APPLICATION NUMBER: PCT/US01/00669

; PRIOR FILING DATE: 2001-01-30

; PRIOR APPLICATION NUMBER: PCT/US01/00665

; PRIOR FILING DATE: 2001-01-30

; PRIOR APPLICATION NUMBER: PCT/US01/00668

; PRIOR FILING DATE: 2001-01-30

; PRIOR APPLICATION NUMBER: PCT/US01/00663

; PRIOR FILING DATE: 2001-01-30

; PRIOR APPLICATION NUMBER: PCT/US01/00662

; PRIOR FILING DATE: 2001-01-30

; PRIOR APPLICATION NUMBER: PCT/US01/00661

; PRIOR FILING DATE: 2001-01-30

; PRIOR APPLICATION NUMBER: PCT/US01/00670

; PRIOR FILING DATE: 2001-01-30

; PRIOR APPLICATION NUMBER: US 60/234,687

; PRIOR FILING DATE: 2000-09-21

; PRIOR APPLICATION NUMBER: US 09/608,408

; PRIOR FILING DATE: 2000-06-30

; PRIOR APPLICATION NUMBER: US 09/774,203

; PRIOR FILING DATE: 2001-01-29

; NUMBER OF SEQ ID NOS: 49117

; SOFTWARE: Anomax Sequence Listing Engine vers. 1.1

; SEQ ID NO 3981

; LENGTH: 1983

; TYPE: DNA

; ORGANISM: Homo sapiens

; FEATURE:

; OTHER INFORMATION: MAP TO AB00658.1

; OTHER INFORMATION: EXPRESSED IN FETAL LIVER, SIGNAL = 8.5

; OTHER INFORMATION: EXPRESSED IN BT474, SIGNAL = 7.1

; OTHER INFORMATION: EXPRESSED IN ADULT LIVER, SIGNAL = 8.3

; OTHER INFORMATION: EXPRESSED IN BRAIN, SIGNAL = 6.3

; OTHER INFORMATION: EXPRESSED IN BONE MARROW, SIGNAL = 7.8

; OTHER INFORMATION: EXPRESSED IN LUNG, SIGNAL = 11

; OTHER INFORMATION: EXPRESSED IN PLACENTA, SIGNAL = 8.3

; OTHER INFORMATION: EXPRESSED IN HELA, SIGNAL = 11

; OTHER INFORMATION: EXPRESSED IN HEART, SIGNAL = 5.6

; OTHER INFORMATION: EXPRESSED IN HEL100, SIGNAL = 15

US-09-864-761-3981

Query Match 3.6%; Score 35.6; DB 9; Length 1983;

Best Local Similarity 52.7%; Pred. No. 0.43;

Matches 77; Conservative 0; Mismatches 69; Indels 0; Gaps 0;

QY 692 ARAAGCCGGTAAAGTTGTTCTACGGCTCAAGAACTCCGCGAGCGCTGTTGGTGGATCG 751

DB 680 AGATTCCAGAAACATCTTCTACTACTGCTGTATGCCCGGTAAGATGGTCTTGATCC 621

QY 752 ATATCGACATCGATGAGGACAAAGCTTGAGTGTCTCCAGGCAGTTACGGAAGACACGGATA 811

DB 620 ACTTTTCCCTCTTTTGGAAATGGCACTGATGTCCACAGGCAGACAGTGTGTGCTCATCTGAA 561

QY 812 GCCTTTGGCAAGGGCCCACTGTTT 837

DB 560 GCACCTAAGTAAATGCCCACTGAGATT 535

RESULT 13

US-10-260-238-3074/c

; Sequence 3074, Application US/10260238

; Publication No. US20040016025A1

; GENERAL INFORMATION:

; APPLICANT: Budworth, Paul R.

; APPLICANT: Moughamer, Todd G.

; APPLICANT: Briggs, Steven P.

; APPLICANT: Cooper, Bret

; APPLICANT: Glazebrook, Jane

; APPLICANT: Goff, Stephen A.

; APPLICANT: Katagiri, Fumiyaki

; APPLICANT: Kreds, Joel

; APPLICANT: Provart, Nicholas

; APPLICANT: Ricke, Darrell

; APPLICANT: Zhu, Tong

; TITLE OF INVENTION: PROMOTERS FOR REGULATION OF PLANT EXPRESSION

; FILE REFERENCE: 6011-NP

; CURRENT APPLICATION NUMBER: US/10/260,238

; CURRENT FILING DATE: 2002-09-26

; PRIOR APPLICATION NUMBER: US 60/325,448

; PRIOR FILING DATE: 2001-09-26

; PRIOR APPLICATION NUMBER: US 60/325,277

; PRIOR FILING DATE: 2001-09-26

; PRIOR APPLICATION NUMBER: US 60/370,620

; PRIOR FILING DATE: 2002-04-04

; NUMBER OF SEQ ID NOS: 6077

; SEQ ID NO 3074

; LENGTH: 1240

; TYPE: DNA

; ORGANISM: Triticum aestivum

; FEATURE:

; NAME/KEY: N region

; LOCATION: (1232)..(1233)

; OTHER INFORMATION: n = any nucleotide

US-10-260-238-3074

Query Match 3.6%; Score 35.4; DB 16; Length 1240;

Best Local Similarity 49.7%; Pred. No. 0.4;

Matches 90; Conservative 0; Mismatches 91; Indels 0; Gaps 0;

QY 750 CGATATCGACATCGATGAGGACAAAGCTTGAGTCTCGTCAGGCAGTTACGGAAGACACCGA 809

DB 344 CGCCCTCGATGATGCTGCCGAAGTGGTGATGCCGACGACCTGGATGAGGGAGAGACTCGA 285

QY 810 TAGCCTTTGGCAAGGGCCCACTGTTTATTTCATCAGTTGTTCGACGAGCCCTGTTGA 869

DB 284 CGGGGGAAGGCGACGCGACGAAGATGGCGTTTCATGGGGTTGTCGGCCATGCCCCGCCGA 225

QY 870 AACCTTACCGGAATACGAAATTTATCTTCCCGGTCCACCGCCTATGTTGGTTCGACGCTACTGT 929

DB 224 CGCCGACTGGTTCAGCAGGAGGTTGTTGAAGGAGCGCAGGTGGTCCGCGGGGCGGA 165

QY 930 C 930

DB 164 C 164

RESULT 14
US-10-229-124-3
; Sequence 3: Application US/10229124
; Publication No. US20030013168A1
; GENERAL INFORMATION:
; APPLICANT: WEI, Ming-Hui et al.
; TITLE OF INVENTION: ISOLATED HUMAN ENZYME PROTEINS, NUCLEIC
; TITLE OF INVENTION: ACID MOLECULES ENCODING HUMAN ENZYME PROTEINS, AND USES
; TITLE OF INVENTION: THEREOF
; FILE REFERENCE: CL001139DIV
; CURRENT APPLICATION NUMBER: US/10/229,124
; CURRENT FILING DATE: 2002-08-28
; NUMBER OF SEQ ID NOS: 5
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 3
; LENGTH: 65042
; TYPE: DNA
; ORGANISM: Human
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (1)..(65042)
; OTHER INFORMATION: n = A,T,C or G
US-10-229-124-3

Query Match 3.5%; Score 34.8; DB 15; Length 65042;
Best Local Similarity 62.8%; Pred. No. 5.2;
Matches 54; Conservative 0; Mismatches 32; Indels 0; Gaps 0;
QY 534 TTTCGAAATAGAAAAGAGCGCTCGTCTATTTTTCGAGGACCAATGGGACATCTTT 593
DB 6894 TTTTGAATAAGAAAAGAGAGTCAACAGTAGAGACTAGACCTCTGGCAATACAAT 6953
QY 594 CTTCCGTCACGAGCGCGCGGAAGA 619
DB 6954 CTTGAGACTAAGACCAGCTGTAAAGA 6979

RESULT 15
US-10-213-878-11
; Sequence 11: Application US/10213878
; Publication No. US20030073206A1
; GENERAL INFORMATION:
; APPLICANT: Bramucci, Michael
; APPLICANT: Nagarajan, Vasantha
; APPLICANT: Thomas, Stuart
; TITLE OF INVENTION: Use of Xylene Monooxygenase for the Oxidation of Substituted
; TITLE OF INVENTION: Monocyclic Aromatic Compounds
; FILE REFERENCE: CL1662 US NA
; CURRENT APPLICATION NUMBER: US/10/213,878
; CURRENT FILING DATE: 2002-08-07
; PRIOR APPLICATION NUMBER: 60/311,490
; PRIOR FILING DATE: 2001-08-10
; NUMBER OF SEQ ID NOS: 22
; SOFTWARE: Microsoft Office 97
; SEQ ID NO 11
; LENGTH: 1038
; TYPE: DNA
; ORGANISM: Sphingomonas sp.
US-10-213-878-11

Query Match 3.5%; Score 34.6; DB 15; Length 1038;
Best Local Similarity 50.2%; Pred. No. 0.69;
Matches 113; Conservative 0; Mismatches 109; Indels 3; Gaps 1;
QY 47 CAGGGAAGAGCCTTTGGTCTCAGCACTTGTCTATGCTATCGGATTTCCGTACAGTGTG 106
DB 77 CGGGAAGACACTTCTTGAAGCGATGCTGGACGCGGGTCTGGCCATGCCGACGATTGCA 136
QY 107 CATCGGAGGTTGGGAGTATGCAAAATTCGAGTTACTCGAAGGGAATGTCCAATCAATGT 166
DB 137 AGTTGGCTCGTGGGTACCTGCAAGTTCAAGCTCGTGTGGCAAGATCGGCAATTGA 196
QY 167 GGCCGGATGCTCCAGGACTTTCTTCGGGAGATCTGTGAAGAGGGCAACCGCCATCTTGCAT 226

Db 197 GCCCGTCGGCCCTTGCACCTTGAGGGCGACGAACCTGCGACGCGCTTTTCGC---CTGCGCT 253
QY 227 GCCAGTGGCTTGGGCTCTCAGACCTCGGATCAAAAGTCGACGTGC 271
DB 254 GCCAGGCCCATTCGCGCTCGGATCTGACAATCGCGGTTGATGCGC 298

Search completed: June 27, 2004, 13:08:52
Job time : 467.449 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: June 27, 2004, 03:11:33 ; Search time 6386.97 Seconds
(without alignments)
10708.573 Million cell updates/sec

Title: US-10-658-691-4
Perfect score: 1578
Sequence: 1 gtcgacccaatgccatt.....ggcaccggctgtgggtag 1578

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 3470272 segs, 21671516995 residues
Total number of hits satisfying chosen parameters: 6940544

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : GenEmbl.*
1: gb_ba.*
2: gb_htg.*
3: gb_in.*
4: gb_om.*
5: gb_ov.*
6: gb_pat.*
7: gb_ph.*
8: gb_pl.*
9: gb_pr.*
10: gb_ro.*
11: gb_sts.*
12: gb_sy.*
13: gb_un.*
14: gb_vi.*
15: em_ba.*
16: em_fun.*
17: em_hum.*
18: em_in.*
19: em_mu.*
20: em_om.*
21: em_ov.*
22: em_or.*
23: em_pat.*
24: em_ph.*
25: em_pl.*
26: em_ro.*
27: em_sts.*
28: em_un.*
29: em_vi.*
30: em_htg_hum.*
31: em_htg_inv.*
32: em_htg_other.*
33: em_htg_mus.*
34: em_htg_pln.*
35: em_htg_rod.*
36: em_htg_man.*
37: em_htg_vrt.*
38: em_sy.*
39: em_htgo_hum.*
40: em_htgo_mus.*
41: em_htgo_other.*

score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB	ID	Description
1	1578	100.0	6442	1	AY026386	AY026386 Rhodococc
C	228.4	14.5	30600	1	AP005935	AP005935 Bradyrhiz
3	224.8	14.2	1521	6	AR041196	AR041196 Sequence
4	224.8	14.2	1521	6	AR116605	AR116605 Sequence
5	224.8	14.2	1521	6	AR159948	AR159948 Sequence
6	224.8	14.2	5215	1	PSEAMNH	D90216 Pseudomonas
7	218.2	13.8	15355	1	AE005026	AE005026 Halobacte
C	208.4	13.2	28050	1	SC093124	AP039124 Streptomy
9	205.8	13.0	310325	1	AE016864	AE016864 Pseudomon
11	200.6	12.7	1958	1	RERAG	D16207 Rhodococcus
12	187.2	11.9	9330	1	AF315580	AF315580 Agrobacte
13	186.4	11.8	4775	6	E12519	E12519 Nucleotide
C	186.4	11.8	208524	1	AE016925	AE016925 Chromobac
15	182.2	11.5	2822	6	BD061400	BD061400 Rhodococc
16	181.8	11.5	10647	1	AE010321	AE010321 Methanopy
17	180.8	11.5	1879	6	AI9015	AI9015 Brevibacter
18	180.8	11.5	2447	1	BLAMDA	M60264 Brevibacter
19	180.8	11.5	3311	1	RSAMD	X54074 Rhodococcus
20	180.8	11.5	7001	1	AE016078	AE016078 Rhodococc
C	180.6	11.4	180124	1	AE017219	AE017219 Geobacter
22	180.4	11.4	7829	1	RER490527	AJ490527 Rhodococc
23	178.6	11.3	11600	1	AE105912	AE105912 Rhodococc
24	175.4	11.1	1566	6	E12517	E12517 DNA encodin
C	174.2	11.0	300700	1	AP006573	AP006573 Gloeobact
25	172.8	11.0	15939	1	AE007128	AE007128 Mycobacte
C	172.8	11.0	318050	1	EX248344	EX248344 Mycobacte
27	170.4	10.8	1482	6	AX863968	AX863968 Sequence
C	170.4	10.8	348676	15	EX842581	EX842581 Mycobacte
30	170.4	10.8	349726	1	EX640421	EX640421 Bordetell
C	168.4	10.7	346259	1	EX640435	EX640435 Bordetell
32	165.6	10.5	10918	1	AE002025	AE002025 Deinococc
33	158.4	10.0	308650	1	AP005218	AP005218 Corynebac
34	156	9.9	1413	6	BD179870	BD179870 Highly th
35	156	9.9	1416	1	AF202447	AF202447 Thermus t
C	155.2	9.8	110000	2	EX255276_03	Continuation (4 of
37	155.2	9.8	110000	2	EX255276_04	Continuation (5 of
38	154.8	9.8	347365	1	EX569691	EX569691 Synchoco
39	154.2	9.8	346287	1	EX640450	EX640450 Bordetell
C	153	9.7	346897	1	AP002395	AP002395 Mesorhizo
41	152.8	9.7	13698	1	AE005329	AE005329 Caulobact
C	152	9.6	215050	1	AL846057	AL846057 Ralstonia
43	151	9.6	17589	1	AE004862	AE004862 Pseudomon
C	146.4	9.3	12115	1	AE014658	AE014658 Bifidobac
C	146.4	9.3	349980	6	AX492782	AX492782 Sequence

ALIGNMENTS

RESULT 1
AY026386
LOCUS
DEFINITION
AY026386
6442 bp DNA linear BCT 12-JUL-2002
Rhodococcus erythropolis ISRhi transposase-like protein gene,
Partial cds; GntR regulator-like protein, enantioselective amidase
(amda), transcription repressor KfrA-like protein, and ISRhl
transposase-like protein genes, complete cds; and unknown gene.
ACCESSION
VERSION
KEYWORDS
SOURCE
ORGANISM
REFERENCE
AY026386.1 GI:21734944
Rhodococcus erythropolis
Rhodococcus erythropolis
Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales;
Corynebacterineae; Nocardiaceae; Rhodococcus.
1 (bases 1 to 6442)

Pred. No. is the number of results predicted by chance to have a

Qy 661 GGCTGTGTCGGTCTTACGGCTTGACATATGACACACCTTGACACATCGGGCC 720
Db 2966 GGCTGTGTCGGTCTTACGGCTTGACATATGACACACCTTGACACATCGGGCC 3025
Qy 721 ATACACAGGGGGTTCAGCTCAACGCCCGGGTCTCTGAGGTGTGGCCGGGGCGACTGG 780
Db 3026 ATACACAGGGGGTTCAGCTCAACGCCCGGGTCTCTGAGGTGTGGCCGGGGCGACTGG 3085
Qy 781 CGGACCTCAGTGGGTGCGTAACCTTCGGAGCGGAGAACTACGGCTCCGGCTCGGC 840
Db 3086 CGGACCTCAGTGGGTGCGTAACCTTCGGAGCGGAGAACTACGGCTCCGGCTCGGC 3145
Qy 841 GAGGAGTATCCGGTCTGAGATTCGGCGTCTCGAGGAGTCACTGGAGCGGAACCGGTGG 900
Db 3146 GAGGAGTATCCGGTCTGAGATTCGGCGTCTCGAGGAGTCACTGGAGCGGAACCGGTGG 3205
Qy 901 ACGCCGAGCTGATCGCGCGGTTCACACGAGACTGCGCGCTTCGAGAGCGCGGTGG 960
Db 3206 ACGCCGAGCTGATCGCGCGGTTCACACGAGACTGCGCGCTTCGAGAGCGCGGTGG 3265
Qy 961 ACCATCGAGGGGTCTCGGTGCGTTCGAGCGCGCTTCGAGAGCGCGGTGG 1020
Db 3266 ACCATCGAGGGGTCTCGGTGCGTTCGAGCGCGCTTCGAGAGCGCGGTGG 3325
Qy 1021 ATGCTTTCACCGCGCGCTATGCGGACTTCGCGGCTTCGCGGTTCGCGGCTTCGCGG 1080
Db 3326 ATGCTTTCACCGCGCGCTATGCGGACTTCGCGGCTTCGCGGCTTCGCGGCTTCGCGG 3385
Qy 1081 CGGTGAGCTCAGACCGCGCTCAGACCGCGCGCTCAGACCGCGCTCAGACCGCGCT 1140
Db 3386 CGGTGAGCTCAGACCGCGCTCAGACCGCGCGCTCAGACCGCGCTCAGACCGCGCT 3445
Qy 1141 CTGCGGCTCTGTCGCGGCTGATGCTGCTGATGCGGAGCACCTGCGGAGCGGCTCCTC 1200
Db 3446 CTGCGGCTCTGTCGCGGCTGATGCTGCTGATGCGGAGCACCTGCGGAGCGGCTCCTC 3505
Qy 1201 GGATTCACCTACCGGAGCGCGAGAACCTCGGCTGAGCTCGGAGCGGCTCGAGCG 1260
Db 3506 GGATTCACCTACCGGAGCGCGAGAACCTCGGCTGAGCTCGGAGCGGCTCGAGCG 3565
Qy 1261 GTCTCCAGGACCGGGTGCATCTGACCCCGACACCGCTACCGTTCGCGGAGCGCTG 1320
Db 3566 GTCTCCAGGACCGGGTGCATCTGACCCCGACACCGCTACCGTTCGCGGAGCGCTG 3625
Qy 1321 TTGAGCGGTGCGGAAGACACATGCTCCATGATCCACGATGACGGGCAATGCGATCCTC 1380
Db 3626 TTGAGCGGTGCGGAAGACACATGCTCCATGATCCACGATGACGGGCAATGCGATCCTC 3685
Qy 1381 AACACGTGCGCGTGGACCTCACCGGTACCGCGCTGACGCTGCCACCGGTCGGGCG 1440
Db 3686 AACACGTGCGCGTGGACCTCACCGGTACCGCGCTGACGCTGCCACCGGTCGGGCG 3745
Qy 1441 GAGAGGGCTGCGCGTGGCTCCAGTATAGGCGCGCTTCGAGGAGTTCGAGCTC 1500
Db 3746 GAGAGGGCTGCGCGTGGCTCCAGTATAGGCGCGCTTCGAGGAGTTCGAGCTC 3805
Qy 1501 TACCGACCGCGCGGTGATGAGCGCGCGCTATGGAGCTCGCGCGGAGCGCGAGC 1560
Db 3806 TACCGACCGCGCGGTGATGAGCGCGCGCTATGGAGCTCGCGCGGAGCGCGAGC 3865
Qy 1561 GCACCGGTGCTGCGGTAG 1578
Db 3866 GCACCGGTGCTGCGGTAG 3883

RESULT 2
AP005935/c 300600 bp DNA linear BCT 12-SEP-2003
LOCUS Bradyrhizobium japonicum USDA 110 DNA, complete genome, section 1/31.
DEFINITION
ACCESSION AP005935 BA000040
VERSION AP005935.1 GI:27348247

KEYWORDS

SOURCE ORGANISM

REFERENCE AUTHORS

TITLE

JOURNAL MEDLINE

PUBMED

REFERENCE AUTHORS

TITLE

JOURNAL MEDLINE

PUBMED

REFERENCE

AUTHORS

TITLE

JOURNAL

FEATURES

source

gene

CDS

gene

CDS

gene

CDS

```
/codon_start=1
/transl_table=11
/protein_id="BAC45268.1"
/db_xref="GI:27348250"
/translation="MVNAGVVVAPPVGMGRHATCGTVGGHIGYRWOIAVSFGLER
QGNADFQGSNANLAFAGVDDSTLDAFLGFTGQGVYANNVLLVVRGAAVVRDKYR
VGFATGLTIDNGSETGWCTVGAGLEFGFAFNWISIGVEYDHLFWGRDRDVFSPGL
VGVVGTAGTATIGQDVLDIGLVRYNRWGSVIATY"
4239..4391
/gene="bsr0004"
/gene="bsr0004"
4391..4391
/gene="bsr0004"
/notes="ORF_ID:bsr0004"
/unknown_protein"
/codon_start=1
/transl_table=11
/protein_id="BAC45269.1"
/db_xref="GI:27348251"
/translation="MGHNVDDVMACVIVAADLCARSACRTFAGADQLGAPRADAKSR
EVDQR"
complement(4427..4750)
/gene="bll0005"
/complement(4427..4750)
/gene="bll0005"
/notes="ORF_ID:bll0005"
/unknown_protein"
/codon_start=1
/transl_table=11
/protein_id="BAC45270.1"
/db_xref="GI:27348252"
/translation="WGLVNHLEVNQRNTATPLSRMSMTATAPLLGLEGITITQS
EYVKGPNVKGHLHCSPVSGKPCSSSFFADFVAADPLAVLGRGFAEVIYVEQTL
AWRLG"
complement(4972..6054)
/gene="nrdB"
/complement(4972..6054)
/gene="nrdB"
/EC_number="1.17.4.1"
/notes="ORF_ID:bll0006"
/codon_start=1
/transl_table=11
/product="ribonucleoside-diphosphate reductase beta
subunit"
/protein_id="BAC45271.1"
/db_xref="GI:27348253"
/translation="MFRMSVIDLSRADPRTLIGKRVGLDSTGSDYVDVRYAWAYEFN
KQQQTHMGEVPLGADLNDWASGRVDSREALTQIFRFTQSDIEVGNYPKRYI
PIFQPLTQVMMAFTNMTETVHIDAYALLKTLGMPKTEFEAFRGYAGRAKADYMEH
FGVDSVADYARTLAFGAFTEGMALFASFAMLLNPPHNKMGMOIYVSWVRDESLH
CEGIKILPHEMRETGAVTRPVRRDDIVDAKTMVLEENFVELAFGLGKIEGMLPEHI
HAYRVADWRLAQLRIAPVFGFEAKESGGFTQLKAHPLPWLVLINGVEHANFEQR
ATEYSKASRGSNDGEDGWATFGRI"
complement(6036..8207)
/gene="bll0007"
/complement(6036..8207)
/gene="bll0007"
/notes="ORF_ID:bll0007"
/probable_ribonucleoside-diphosphate reductase (EC
1.17.4.1)"
/codon_start=1
/transl_table=11
/protein_id="BAC45272.1"
/db_xref="GI:27348254"
/translation="MLGNPLCRQLFRGISATLGRPTAAATREPGNLPRONPCPRA
GCVGRLOTFARSPKRSARSOPNITGLPMSASYDFDOYGHLELPSVROPSCQA
KPTTELQVQPPALPRLPPRPSPTPSDLSDRSDDLTSFGRLTLTDVLLAG
EGLQMFARVSCAFADNVAHQRLYDMSRLWFNPATFVLSNGTTLGLPLSCFNSV
SLSLGI VGTNENWNLASGGIGTYGNVRSIGKARGVTSIGIIPFHTVDGLTL
ALQSGLRGSAAVYLDLHPEIEEFLIRKASGDFNRKSLNHHGINVTDFNCVAG
ACTALPSPKTNVREYIDARLQWRIENRLQTEPYLLFIDTVNRLPKHQRELQ
LEVSTNLSCEITLSTGIDHREERTAVCCSLNLETWQDNEEPGFIEDVMEFLDN
VLTHITVAPQMKARVAALRERSVGLGVMGHFSFLOAKGI PMEGTLAKSFNLMFR
KLRRADAASVALARGPCPDALRGLRWARFESHKIAIAPTASISICGTSACTEPI
```

```
gene
CDS
```

```
PANIYTHKLSGALSVRNPHLAKVLAAGKADTSATWESTIAHDSGVAHLDVLSHEHKA
IFRTAFEDQRWIVELAADRAALICQSINLYIPADVDKMDLHMLHWTAMKRGVKSLL
YYCRSISRAAPAGKIAGDATPTQPARQRTDYERCLACQ"
complement(8380..8688)
/gene="bll0008"
/complement(8380..8688)
/gene="bll0008"
/notes="ORF_ID:bll0008"
/unknown_protein"
/codon_start=1
/transl_table=11
/protein_id="BAC45273.1"
/db_xref="GI:27348255"
/translation="MARRYWLKPDPAQLNSRTANRSMPTFRMSRVAVLLENLRHFP
ORFHLADRRRVVVERTFANLGRCHLCKDFECSAATLAWPLVAHRLRLRLARP"
complement(8962..9714)
/gene="bll0009"
/complement(8962..9714)
/gene="bll0009"
/notes="ORF_ID:bll0009"
/putative_acetyltransferase"
/codon_start=1
/transl_table=11
/protein_id="BAC45274.1"
/db_xref="GI:27348256"
/translation="MHPASPRRCIHSARQMCRCCHAPPKWKANAKTEPRNDFMTEBDSI
RSANANIRASTVWTIVLGRADPLNWSWPDPROVLRSMPOFIDACGAFAGATYAV
AGIRGALWLPVNVQDDEALDAINALSIRFEITEDMAHLRRMAEHRPHEHWILP
LIAVDNWKVGKGLTLMNYAUCRDEGDGIVAYLESSNPENPFYQRHGKGVKVGIEIQH
GSPPPTLMRLTASFRPSKAAPDRLVTQQA"
complement(10345..10743)
/gene="bll0010"
/complement(10345..10743)
/gene="bll0010"
/notes="ORF_ID:bll0010"
/unknown_protein"
/codon_start=1
/transl_table=11
/protein_id="BAC45275.1"
/db_xref="GI:27348257"
/translation="MLPRLQFSKTERNTHVLDLPGCVASIPAPRLNPNRSDQGV
WNSYBAINACQKPLLRLLQLRQPAVFPFMARIAATTLDLGVYDEGTTATFSLD
RANAVRAIDTDDQSMHSSRAIPSCVVQ"
complement(10872..11492)
/gene="bll0011"
/complement(10872..11492)
/gene="bll0011"
/notes="ORF_ID:bll0011"
Query Match 14.5%; Score 228.4; DB 1; Length 300600;
Best Local Similarity 50.9%; Pred. NO. 4.1e-21;
Matches 677; Conservative 0; Mismatches 636; Indels 17; Gaps 5;
Qy 217 CCGACGGCGGACGAAGACCCGTTCAACGGCATTCATCCGGTTCTTCCGGGTGAGGCGCGC 276
Db 194385 CCCTCGGCGAAGAGAACCCGCGCATGCTGTATCGCAAGCGCATGTGAAGCGCC 19436
Qy 277 ACGAGAGGGCCATGAGCGACCTGACCGCGCGGATCAAGACTGATGCCCATCGCCGGT 336
Db 194325 GCCGCGCGCAAGCTCAAGGGCAAAACCTTGCGCTGAAAGACAACATCATGTCGCCCGC 194266
Qy 337 ATGCCACACCAAGAAACGGGTCCCGGATGCTCCCGACTGTATGCCACCGAGGATGCCGTG 396
Db 194265 GTGCCCATGACCAACCGGCTCGTCGAGCTGGAAGGCTATGTGCCCATTCGACGCCACC 194206
Qy 397 GTGGTGAAGCGGTGTCTCGCGGAGCGGCAACCATCGTCGGAAGACGAACTCGAGGAC 456
Db 194205 ATCGTCACGGCATGTGGATGCCGTGCCGATCAAGGGCAAGGTGATCGCAGCAT 194146
Qy 457 ATGCGCATG---GGTATCGGTGAAGCAGCGTCTACGGTCTCGCGCTGACCGCAAC 513
Db 194145 TTCTGCTGTCCGCGCGGACGACCCGGCTCTTTTCGGGCCCGTGACAAATCCGCAAG 194086
```

```
gene
CDS
```



```
QY 514 CCGCCACGCGCAGCGGTGATCTCCAGCGGCTCCGGGCTCCCGTCTGCGCGCATS 573
Db 194085 ATGGGTACTCGCGCGGTGCTGCTCAGGCTCGGGCTCGTGTCTGCTCGCGGAA 194026
QY 574 GTGAGCTTCGCGCTCGGCGTGCATGAGCGAGCAGTCCGATCCGCGCGCATGCTG 633
Db 194025 GTGACATGCGCATCGCGCGCATGAGGGGCTCGATCCGATCGCGCTCTGCTCTGC 193966
QY 634 GGAAGTGTGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 693
Db 193965 GGCATTTACGGCATGAAGCGAGCTGGGCGCTCGTCTTACACCGGATCATGCCGATC 193906
QY 694 GACCAACCTTGACACCATCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 753
Db 193905 GAGATCTATGTCAGCACATACCGGTCCGATACCGCGCGCGCGCGCGCGCGCG 193846
QY 754 CTGAGGTGTGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 813
Db 193845 CTTGAAGTGTGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 193792
QY 814 CCGGAGAACTACGGCTCCGGGCTCGCGGAGGAGTATCCGCTGAGATTCGGGCTG 873
Db 193791 GTCAGGAATACACCAAGGCGCTCGCGGAGGCTCAAGGCGATGAGATTCG 193732
QY 874 GAGGAGTCACTGGAGCGGAGCGGTCGACCGCGCGCGCGCGCGCGCGCGCG 933
Db 193731 AAGGAAGCTTTGAGCAGCGGTCGACAGAGCGCGGTGATGAAGCGTGCAGGG 193672
QY 934 CTGGCGCGCTCGAGAGCGCGGTCGACCATCGAGCGGCTCGGTCGCGGTGTCGAG 993
Db 193671 GCAAAGCGCTTCAAGGATCTTGGCGCGCGCGCGCGCGCGCGCGCGCG 193612
QY 994 CGCGCTTGGCTATTCAGAGCGCGGTCGATGCTTTCAACCGCGCGCGCGCG 1053
Db 193611 CTGGCGCGCGGATCTGACCGCGATCGGACCGCGCGCGCGCGCGCGCGCG 193552
QY 1054 GCGGCTGCGGCTACTTCCAAAGGCGCGGTCGAGCTCAGACCGCGCGCGCG 1113
Db 193551 GCGACGCTACGCGCTCAGCGCGCGGATCTCTATTTCGACCTCGCTGATG 193492
QY 1114 GCGGCTGCGGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 1173
Db 193491 CGCGCTGCGGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 193432
QY 1174 GCGGAGCACCTGCGCGCGCGGATCTCTCGGCTACCTACGCGAGGCGCGAG 1233
Db 193431 GGCACCTACATCAACAATTTCCGTCACCGCTTCTACGCGAGGCGCTCAAT 193372
QY 1234 CTGAGCTCGGACGAGCATGACCGCTCTCAGGACCGGCTGACCTGACCGCG 1293
Db 193371 GCGAGGCTGACCGCGCGCTACGAAAGCGGTTGCGGATTAAGATCTGCTG 193312
QY 1294 ACCACGCTACCGTTGCCAAGAGCTG-TTGAGCGGTCCGCAAGACACCATG 1352
Db 193311 ACAGCGCGATGAAGGCAACCAAGCTCGGAGCGCCACGCGCGCGGAGGACT 193252
QY 1353 CCCACGATGACGGGCAATGCGATCTCAACAGTGGCGCTGACCTACCGGTACCC 1412
Db 193251 GCGCGTG---CGTGGAGATGATCGGCAACACCGCGCGGTTGACATCAACCA 193196
QY 1413 GCGGCTGACGCTGCCAAGGCTGCGGCGGAGAGGCGCTGCGGCTTCCAAAGT 1472
Db 193195 CGGATGTCGCTGCGGCTGCGGCTGCGGCTGCGGCTGCGGCTGCGGCTG 193139
QY 1473 AGCGCGCACCTTCAGAGGATGAGCGCTTACCGACCGGCGCGGCTGATTCAGG 1532
Db 193138 CGCGCGCATGTCAGAGGATGAGCATCTATCGCGCGCGCGCGCGCGCGCGCG 193079
QY 1533 CCTATGGGAG 1542
Db 193078 CCACTGGGAG 193069
```

```
RESULT 3
AR041196
LOCUS AR041196 1521 bp DNA linear PAT 29-SEP-1999
DEFINITION Sequence 20 from patent US 5811286.
ACCESSION AR041196
VERSION AR041196.1 GI:5961692
KEYWORDS
SOURCE Unknown.
ORGANISM Unknown.
REFERENCE 1 (bases 1 to 1521)
AUTHORS Fallon,R.Donald., Nelson,M.James, and Payne,M.Scott.
TITLE Nucleic acid fragments encoding stereospecific nitrile hydratase
and amidase enzymes and recombinant organisms expressing those
enzymes useful for the production of chiral amides and acids
JOURNAL Patent: US 5811286-A 20 22-SEP-1998;
FEATURES Location/Qualifiers
source 1..1521
/organism="unknown"
/mol_type="unassigned DNA"
```

Query Match 14.2%; Score 224.8; DB 6; Length 1521;
Best Local Similarity 48.8%; Pred. No. 5.5e-20;
Matches 731; Conservative 0; Mismatches 752; Indels 15; Gaps 4;

QY 37 CCGACAGCGGACAGCTCCAGAGGTACAGCGCGCGCCACCACTTCGACCTCGACGAGAA 96
Db 16 CTTACCTCTGACCGAGTTTGTAGACATCCGAACCCAGTTGCACTCACTACGACCGAA 75
QY 97 CTGGCGCGCGCAGCTCGTTCCGCTCGTGGCGGAGATGGTGACTGCTTCCACTGATCGAC 156
Db 76 CAGGCGAGCGTCTACTGGAAGTATGCAACCGAGTTTCGACGCGCTACGACTGCTGCGAC 135
QY 157 GAACTACGCGAACCGCGCGCGCGCGCGGTACAGCGCGCGCGGACATCGCGCGGAA 216
Db 136 GAACTGGCTGATTTTCGTCGCGCAATACGTAAGACCGAGTTTCAGGCTATCGCCATCG 195
QY 217 CCGACCGGCGAGAGAGACCGTTCAACGCACTTCATCCGCTTCTGCGGGTGGAGGCGCC 276
Db 196 CCATCGGCGAAGAAACCCCTCTGAACGCTGGTACTACCGAAACAGAAAGTGAATGGTGC 255
QY 277 ACGGAGGCGCACTGAGCGACTGACCGCGCGCGGATCAAGGACTGATCGCATCGCCGCT 336
Db 256 CCGGAGGCGCTCTGCGCGGCGAAACCGCTCGCGCTCAAGATAATATCTCCCTGGCAGGC 315
QY 337 ATGCCACACCAAGCGGTCCCGATGCTCCCGATGTCGCGCACCGGAGGATCGCGGT 396
Db 316 GTCCCATGATGAACGCGCGCGCGCTTGGAAAGCTTCGTCGCGGGTTCGATGCCACG 375
QY 397 GTGGTGGAGCGCTGCTCGCGCGAGCGCGCACCATCTGTCGGCAAGACGAACTCGAGGAC 456
Db 376 GTGGTCAACCGCTTCTCGATCGGGGCGGACCAATCTCGGCAAGCGCACCTCGCGAGC 435
QY 457 ATGGCGAT---GGGTATCGTGAAGCGAGCGCTCTACGCTCGCTCGCTGAACCGCAAC 513
Db 436 TACTGCTTTCAGGAGGAGCGACACCTCCGATCAGCGCGCGGTGCAACCCACATCGC 495
QY 514 CCGCGCGCGCGCGCGGTGATCTTCAGCGGCTCCGCGGTGCGGTGCGGTGCGGTGCG 573
Db 496 CAGGTTATGCTCTGCGGCTTCTCATCAGGCGCGCGGCTATGCTGCTGCGGTGAG 555
QY 574 GTCGATTCGCGCTCGGCGTTCGATGAGCGAGCGACATCCGATCCGCGCGCATGCTGC 633
Db 556 GTGGACATCGCGTGGCGGCGATCAAGCGGCTCATTCGATCCGCTCGGCTTCTGC 615
QY 634 GGAAGTGTGCGGATGAAGCGCGCGCGCGCTGCTGCGCTTTCAGGCTGATGATG 693
Db 616 GTTACCTACGCGATGAAGCGCGCGCGCGCTGCTGCGCTTTCAGGCTGATG 675
QY 694 GACCAACCTTCGAGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGGTC 753

Db 676 GAAGCAGATCGATCATGTGCGGCCCATACCGGTAAAGTGGCGGCAACACGGCTGATG 735
QY 754 CTCGAGGTGTTGGCCGGGCGGAGTGGCGGACCCCTCAAGTGGGTGCGTAACTTCCGGAG 813
Db 736 CTCGAGGCAATGGCCGGTGCAGACGACTCGACCCGCGCCAGCGCGCG-CTTCAG 789
QY 814 CCGGAGACTACGGCTCCCGCTGGCGGAGGAGTATCGGTCTGAGATTGCGGTGTC 873
Db 790 GTGATGACTATTGAGTTACTTGGAAAAGCGGTGAGCGGACTCAGAACTCGGGTGTG 849
QY 874 GAGGAGTCACTGAGCGGCAACGGTGCAGCGCGGACGTGATCGCGCGCTTCAACACAGGA 933
Db 850 CAAGAGGATTCGCGCTTCTAACAGGACCTCGCGTGGGGAACAAAGTGGCGACCC 909
QY 934 CTGGCGGCTCGAGAGCCGGTGGCGGACCTCGAGCGGCTCGGTGCGGTGAGG 993
Db 910 ATGCGCGGACTCGAGCGGTGGCGCTCATGTGAGCGGCTCTCATTCGCGAGCAAC 969
QY 994 GCGGCTGCGCTATCCAGAGCGCGCTGATGGCTTTCAACGCGCGCGCTATGCGCGACTCC 1053
Db 970 CTGGCAGGCTGTGTGGCACCCCATCGGTTCGGAAGGCTTGACCATGAGATGATG 1029
QY 1054 GCGGTGTGGGTACTTCCAAAGGGGCGGTGGAGCTGAGCAGCGCGCTGACAGCGGG 1113
Db 1030 GGCACGCGCGAGGCTTTAACTGGAAAGGACTTACGATGTGCGGCTGCTGACAAACAA 1089
QY 1114 GCGGAGTGGACACCAACGATCTGGCGATCTGGCGGCTGATGCTGTGATC 1173
Db 1090 GCGAGTGGCGGACGACGACGACCAATATCGCGTGCCTCAAGCTCTGATGTGTC 1149
QY 1174 GCGGAGACTCTGGCGGACGAAATCTCTGGGATCCACTACGGAAGCGCGAGAACTTCGG 1233
Db 1150 GGCACATAGCGCTGTGGGCTGACAAAGGAGCTTACGATGTGCGGCTGCTGACAAACAA 1209
QY 1234 CTGGAGTGGGAGCAGATGACGCGCTCTCCAGGACCGGCTGCTGCTGACCGCG 1293
Db 1210 CGCTTTGCGCGGAGGATGACGAAAGCGCTGCAACCTATGACCTGTGTGATGCG 1269
QY 1294 ACCAGCTACCTGTGCGGACGAGTGTGAGCGGTGCGGAGACACCAATGTCATGATC 1353
Db 1270 ACCAGCTACCTGCGGCGGACGCGGCTGCGGCGGCTGCGGCTGCGGCTGCGGCTGCG 1326
QY 1354 CCAGGATGAGGCGATGATCTCAACAGTGGCGGCTGCGGCTGCGGCTGCGGCTGCGG 1413
Db 1327 GTGGCTGCGGCTGGAATGATGCGGATACCGGCGGACAGCATACCGGCGGATCGG 1386
QY 1414 GCGCTGAGCGTGGCGGCGGAGGAGGCGGCTGCGGCTGCGGCTGCGGCTGCGGCTGCG 1473
Db 1387 GCGATGTCGATTC- -CGTGTGCGCTGCTGAGCGGCTGCGGCTGCGGCTGCGGCTGCG 1443
QY 1474 GCGGCGGCTGCGGAGGCTGAGCTTACCGGACCGGCGGCTGATGAGGCGGCG 1531
Db 1444 GCAAAACACTACGCGGAGGCGATTTACCAAGCGGCGGCGGCTGAGGCGGCTCGG 1501

RESULT 4
ARL16605
LOCUS ARL16605 1521 bp DNA linear PAT 16-MAY-2001
DEFINITION Sequence 20 from patent US 6133421.
ACCESSION ARL16605
VERSION ARL16605.1 GI:14096927
KEYWORDS
SOURCE Unknown.
ORGANISM Unknown.
REFERENCE 1 (bases 1 to 1521)
AUTHORS Fallon,R.Donald., Nelson,M.James. and Payne,M.Scott.
TITLE Polypeptides and polypeptide subunits of a stereospecific nitrile
hydratase enzyme
JOURNAL Patent: US 6133421-A 20 17-OCT-2000;
FEATURES Location/Qualifiers
source 1..1521
/organism="unknown"

ORIGIN
Query Match 14.2%; Score 224.8; DB 6; Length 1521;
Best Local Similarity 48.8%; Pred. No. 5.5e-20;
Matches 731; Conservative 0; Mismatches 752; Indels 15; Gaps 4;
/mol_type="unassigned DNA"
QY 37 CCGACAGCGAAGCAGCTCCAGGAGTACAGCGCGCGCCGACACTTCGACCTCGAGCAGGAA 96
Db 16 CCTACCTTCAGCAGGTTTTCAGATCCGAAACCGATTCGACATGCAATGACGACGAA 75
QY 97 CTGGCGGCGGCTCGTTCGGTTCGTTGGCGGAGATGTTGCTTCGACCTTCGACCTGATCGAC 156
Db 76 CAGGAGCGTCTTACCTGGAATGATGCAACCGAGTTTCGACGCTTACGACCTGTCGAC 135
QY 157 GAATACCGCAACCGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 216
Db 136 GAATGCTGATTTTCGTTTCGCGCAATGACGCTAGCAGCTTACGCTATCGCATCGG 195
QY 217 CCGACCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 276
Db 196 CCATCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCG 255
QY 277 ACGGAGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 336
Db 256 CGGAGGCGGCTGTCGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCG 315
QY 337 ATGCCCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 396
Db 316 GTCCCGATGATGAAAGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 375
QY 397 GTGGTGAAGCGGCTGTCGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 456
Db 376 GTGGTCAACCGGCTGTCGATGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 435
QY 457 ATGCGGAT- -GGGTATGCGGTGAAGGCGGCTGATCGGCTGCTGCGGTGAACCGGCGG 513
Db 436 TACTGCTTTCAGGAGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 495
QY 514 CCGCGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 573
Db 496 CACGCTTATGCTTGGCGGCTTCTCATGAGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 555
QY 574 GTCCAGCTTCCGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 633
Db 556 GTGGAGTTCGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCG 615
QY 634 GGAAGTTCGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 693
Db 616 GGTACCTAGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 675
QY 694 GACCACTTGGAGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 753
Db 676 GAAGCGGATGATGATGTCGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCG 735
QY 754 CTCGAGGTGTTGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 813
Db 736 CTGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 789
QY 814 CCGGAGAACTAGCGGCTCGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 873
Db 790 GTGATGACTATTGAGTTACTTGGAAAAGCGGTGAGCGGACTCAGAACTCGGGTGTG 849
QY 874 GAGGAGTCACTGAGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 933
Db 850 CAAGAGGATTCGCGCTTTCGTTTACAGGACCTTCGCGTGGCGGCAAAAGTGGCGGCGG 909
QY 934 CTGGCGGCGGCTCGAGAGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 993
Db 910 ATGCGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 969
QY 994 GCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 1053

Db 970 CTGGCAGGTTGTGTGGCACCCTCTGAAAGGCTTACCATGTCAGATGATGCAT 1029
QY 1054 GCCGGTGTGGGCTACTTCCACAAAGGCGCGTGGAGCTCAGCAGCCGCTCAGCGCGG 1113
Db 1030 GGCAACGGCGCAGGCTTTACTGGAAGGACTTACGATGTGGCTGTGACAAACA 1089
QY 1114 GCCAGAGTCGACCCACCAAGGATCTGGGATCTCTTCCGGGTGATGTGTGATC 1173
Db 1090 GCCAGTGGCGGACGACGAGCAACCAATATCCGCTGCTCAAGCTCTGCTGTCGTC 1149
QY 1174 GCGAGCCTCGCGCAGCAATACCTCGGATCCACTACGGAAGCGCAGAACCTGCGG 1233
Db 1150 GGCAATACGGCTGTGGGCTACAGCGCTACTACGCAAGGCCAGAACCTTGCA 1209
QY 1234 CTGGAGCTCGCAAGCAGATCGACCGCTCTCCAGGACCGGCTGCTACCGCCG 1293
Db 1210 CGCTTTGCCCGGAGGATACGAAAGCGCTGCAAACTATGACCTGTGGTGTGTCG 1269
QY 1294 ACCACGCTACCTTCCCAAGGCTGTGAGGCTCGGCAAGCAGACCATGTCATGTC 1353
Db 1270 ACCACGCTACCTTCCCAAGGCTGTGAGGCTCGGCAAGCAGACCATGTCATGTC 1326
QY 1354 CCAGGATGACGGGCAATCGATCTCAACAGCTGCGGCTGACCTACCGCTCACCGG 1413
Db 1327 GTGGCTCGCGCTTGGAAATGATCGGCAATACCGGCCACAGACATCACCGGCTCG 1386
QY 1414 GCGCTGACGCTGCGCAGGCTGCGGCGGAGAGGCTGCGGCTTGGCTTCAAGTATA 1473
Db 1387 GCGATGTGATTC---CGTGTGCTGCTGAGGCGCTGCGGCTGATGTGCTGTC 1443
QY 1474 GCGCGGCTGCTGAGGAGTCTGAGCTTACCGGACCGGCGGCTGATGAGCGCGCG 1531
Db 1444 GCAAAACACTACCGGAGGCGAGATTACCAAGCGCGGCGGCTTGAAGCTCGG 1501

RESULT 5
AR159948
LOCUS AR159948 1521 bp DNA linear PAT 17-OCT-2001
DEFINITION Sequence 20 from patent US 6251650.
ACCESSION AR159948
VERSION AR159948.1 GI:16222813
KEYWORDS
SOURCE Unknown.
ORGANISM Unknown.
REFERENCE 1 (bases 1 to 1521)
AUTHORS Fallon,R.Donald., Nelson,M.James, and Payne,M.Scott.
TITLE Pseudomonas putida amidease polypeptide useful for the production of
chiral amides and acids
JOURNAL Patent: US 6251650-A 20 26-JUN-2001;
FEATURES
source
1. .1521
/organism="unknown"
/mol_type="unassigned DNA"

ORIGIN
Query Match 14.2%; Score 224.8; DB 6; Length 1521;
Best Local Similarity 48.8%; Pred. No. 5.5e-20;
Matches 731; Conservative 0; Mismatches 752; Indels 15; Gaps 4;
QY 37 CCACAGCGGACAGCTCCAGGAGTACAGCGCGCCGACCTTCACCTCGACGAGAA 96
Db 16 CCTACCTCGACGAGTTTATGACATCCGAACCACTTGCATGCACTGACGACGAA 75
QY 97 CTGGCGCCGACGCTGCTCGGTCGTCGGGAGATGCTGCTTCCGCTGATCGAC 156
Db 76 CAGGACGCTCTTACCTGGAATGATGCAACCGAGTTTCGACGCTTACGCTGCTG 135
QY 157 GAACTACCGAACCGCGGAGCGCGGCTTACGAGCGCGGCTGCTGCTGCTG 216
Db 136 GAACTGCTGATTTGTTTGGCCCAATACGCTACGACCGGAGTTTACGCTGCTG 195
QY 217 CCAGCGGCGGAGAACCGGCTTCAACGCTTATCCGCTTCTCGCGGCTGGAGGCGCC 276

196 CCATCGGCAAGGAAACCCCTCTGAAAGGCTTACCGAAACAGAAAGTGAATGGTGCC 255
277 ACAGAGGGGCTACGAGGACCTGACCGCGGCTGAGGACTGCAATGCGCATGCCCGGT 336
256 CGCAAGGCTGCTGCGGCGGCAAAACCGTCCGCTCAAGATATATCTCCCTCGCAGGC 315
337 ATGCCCAACCAAGGCTTCCGCGATCTCCGACTGTGATCGCAACCGAGGATGCCGTG 396
316 GTCCCATGATGAACGCGGAGCGCGTGGAGGCTTCGTCCCGGCTTCGATGCCACG 375
397 GTGCTGAGCGGCTGCTCGCGGAGGCGGACCACTGTCGCAAGAGCAAGCTCGAGAC 456
376 GTGCTACCGGCTTGTGCTGATGCGGCGGCGACCAATCTCGGCAAGCAACCTCGGAC 435
457 ATGGCGAT---GGGTATCGGTGAAGGAGGCTTACGCTCTGCGCTGAACCCGAAAC 513
436 TACTGCCTTTTCAGAGGAGCAGCACCTCCGATCCAGCCCGGTGCAACACCATCGC 495
514 CCAGCCCAAGGCTACGCTGATCTTCAGCGGCTTCGCGGCTGCGCTGCTCGCGCATG 573
496 CACGTTTATGCTTCTGCGGTTCTCATAGGAGCGCGCATTTGGTTGCTCGGCTGAG 555
574 GTGACTTTCGCGGCTGCTGATGAGGAGGAGCATCCGATCCCGGCGCATGCTGTC 633
556 GTGACATCGCTGCGGCGGCTCAAGGCGCTCAATCGATCCGCTGCGCTTCTGC 615
634 GGAATGTGCGCATGAAGGAGCAACCGGCTGCTGCTTACGCGCTGACATACATG 693
616 GGTACTACGCGATGAAGGCGCACCCCGGCTGCTGCTTACACCGGCTCATGCGCAT 675
694 GACCAACCTTGAACCATCGGCGCATCACAGGCGGCTGAGCTCAAGCCCGGCTG 753
676 GAAGCAGATCATGTCGCGCCCATCACCGTAACTGCGGCAACAGCGCTGATG 735
754 CTCAGGTTGTGCGCGGCGCATGTCGCGCAACCTCAGTGGGTGCTAACTTCGCGAG 813
736 CTGAGGCAATGCGGCTGCGAGCGGCTGCAACCGCGCGGCGGCTGCTGCTGCTG 789
814 CCGGAGGACTACGCTCGGCTGCGGAGGAGTATCCGCTGCTGAGTTCGCGCTGCTG 873
790 GTCGATGACTTATGCTGCTTACCTGGAAGGCGGTGAGCGGCTCAGAACTCGGCTG 849
874 GAGGATGCTGAGGCGGAGCGGTGCGACCGCGGCTGATGTCGCGGCTTCAACAGGGA 933
850 CAAGAGGATTCGCGCTTGTCTAACAGGACCTTCGCTGCGGCAAAAGTGCAGCGCC 909
934 CTGGCGGCTGCGAGGCGGCTGCGGACCATGAGCGGCTGCTGCTGCTGCTGCTG 993
910 ATCGCGGCTGCGAGGCGGCTTGGCGCTCATGTCGAGCGGCTTCTCATTCGCGAGCA 969
994 GCGGCTGCGGCTATCCAGAGCGGCTGATGCTTCAACCGCGCGGCTATGCGGAGCTCC 1053
970 CTGGCAGGTTGTTGTGCAACCCCATCGGTTGCAAGGCTTGACATGCAATGATGCAT 1029
1054 GCGGCTGCTGCTGCTTCCCAAGGCGGCTGAGCTGAGCAGCGCGCTCAGACGCGG 1113
1030 GCGAACGCGCGGCTTAACTGGAAGGAGCTTTACGATGTCGCGCTGCTGCGCAACAA 1089
1114 GCCCAGAGTGCACACCAAGGATCTGCGGATCTGCTCCGCGCTGATGCTGCTGATC 1173
1090 GCGAGCTGCGCGGAGCGGAGCAATTTATCCGCTGCTGCTGCTGCTGCTGCTGCTG 1149
1174 GCGAGGACCTGCGGAGCAATACCTCGGATCCACTACGGAAGGCGGAGAACTCGGG 1233
1150 GCGCAATACGCGCTGCTCGGCTACAGGAGCGCTTACTACGCAAGGCGGCGGAGAACT 1209
1234 CTGGAGCTCGCAAGCAGATCGACCGCTCTCCAGGACCGGCTGCTGCTGCTGCTG 1293
1210 CGCTTTGCCCGGAGGATACGAAAGCGCTGCAAACTATGACCTGCTGCTGCTGCTG 1269
1294 ACCAGCGCTACCGTTCGCAAGGAGCTGTGAGCGGCTGCGGAGCAACCATGCTCATG 1353

```

Db 1270 ACCACGCCCATCAGCGCCCAACCCACCCGCCAG---CGAACTGCTCGATCACGGAGTAC 1326
Qy 1354 CCACGGATGACGGGAATGCGATCTCAACAGCTGCCGCTGGAGCTTCCACGGGTCAACCG 1413
Db 1327 GTGGCTCGCGCTTGAATGATGCGGATACCGGCCACAGGACATCACCGGCGATCCG 1386
Qy 1414 GCGTGACGTGCCACCGCGTGGCGGCGAGAGGCGCTGCCGCTTGGCGCTCCAGTGAATA 1473
Db 1387 GCGATGTCGATTC---CGTGTGGCTGCTGGACGCGCTGCCGCTGGGCTGATGCTGTC 1443
Qy 1474 GCGCGCCACTTCGAGGAGTCGACGCTCTACGCGACCGCGCGGCTGATCGAGCGCGCG 1531
Db 1444 GCARAACTAGCCGCGAGGACGATTACCAAGCGCGCGCGGCTTTGAAGCTCGG 1501

RESULT 6
LOCUS PSEAMNH
DEFINITION Pseudomonas chlororaphis 5215 bp DNA linear BCT 21-DEC-2002
hydrate alpha and beta subunits.
ACCESSION D90216 M65204
VERSION 91193202
KEYWORDS D90216.1 GI:216850
SOURCE Pseudomonas chlororaphis
ORGANISM Pseudomonas chlororaphis
Bacteria; Proteobacteria; Gammaproteobacteria; Pseudomonadales;
Pseudomonadaceae; Pseudomonas.
REFERENCE 1 (bases 1 to 5215)
AUTHORS Nishiyama,M., Horinouchi,S., Kobayashi,M., Nagasawa,T., Yamada,H.
and Beppu,T.
TITLE Cloning and characterization of genes responsible for metabolism of
nitrite compounds from Pseudomonas chlororaphis B23
JOURNAL J. Bacteriol. 173 (8), 2465-2472 (1991)
MEDLINE 91193202
PubMed 2013568
COMMENT These data kindly submitted in computer readable form by: Makoto
Nishiyama
Department of Agricultural Chemistry
The University of Tokyo
1-1-1 Yayoi, Bunkyo-ku
Tokyo 113
Japan
Phone: 81-3-3812-2111 x5126
Fax: 81-3-3812-0544.
Location/Qualifiers
1. =5215
/organism="Pseudomonas chlororaphis"
/mol_type="genomic DNA"
/strain="B23"
/db_xref="taxon:333"
245. .248
/notes="SD sequence for amidase"
254. .1774
/EC_number="3.5.1.4"
/codon_start=1
/transl_table=11
/product="amidase"
/db_xref="GI:216851"
/translation="MAITRTDQVLDIRTLHMQLTAEQASYLELMQPSFDAYDLV
DLGVPMNGAAPLEGVPGFDATVTVRLDAGATILGKATCEHYCLSGHSTSPAPV
HNPHRGVASGSSGSAALVAGSEVDIAVGDDGGSIRIPSAFCGYGKPYGLVP
YTGVAIETADIHVPIGTVNRDNLALQAMAGADGLDPROAQPQVDYCSLEKGV
GLRIGLQGFALANQDPRVADKVRDAIARLEAGAHVPEVSIPEHNLGLLHPIC
EGLTQMHNAGFNWGLXVGLLDQKASWRDDQLSASLKLCMFVGYGSLRYN
GRYAKAQLARFAQWGLKALQTYLDLVMPTTITAQHPFANCSTIYVARALEMI
GNTAQDITGHFAMSIPOGLLDGLPVLGMLVAKHYAEGTIYQAAAAFEASVDWRTL
1857. .1861
/notes="SD sequence for alpha subunit of nitrite hydratase"
1869. .2471
/EC_number="4.2.1.84"
/codon_start=1

```

```

/transl_table=11
/product="nitrite hydratase alpha subunit"
/protein_id="BAA14245.1"
/db_xref="GI:216852"
/translation="MSTSIATSTATPGERAWALFOVLKXELIPEGVTEQLTQIMA
HWSPENGARVAKAWVDVDFRALLKKGDTACACQFGYTPQGEITVALEDTPGVKNV
IYVCSLCSPLNFWLGLPPEWYKGFPRALRVEGRVLRBELTLPSTVIKWDVTS
ESRYLVLPORPEGSEHMBEQQLVTKDVLIGALPRVG"
2502. .2505
/notes="SD sequence for beta subunit of nitrite hydratase"
2514. .3176
/EC_number="4.2.1.84"
/codon_start=1
/transl_table=11
/product="nitrite hydratase beta subunit"
/protein_id="BAA14246.1"
/db_xref="GI:216853"
/translation="MDGFHDLGFGQFGKVPHTINSYKQVFKODWEHLAYSIMFVG
VPLKKFSDVEYHAEVLDPQHYGTQYXYEYIATATLAVETGVIQAEIDQALGS
HPKLANPAHATGRPAITGRPPPEGVDRVVRDEYVAGHIRMPAIYVRGEGVYHRTSE
QMPFPDAIGHGDLSSAAHQPTYHVEPRVDLWGDAAADGGVYVVDLFEYLKAPGAQV
NA"
3160. .3168
/notes="SD sequence for P47K"
3173. .4432
/notes="unnamed protein product; P47K"
/codon_start=1
/transl_table=11
/protein_id="BAA14247.1"
/db_xref="GI:216854"
/translation="MIEGAQAGRLPVTVLSGLFAGKTKLLNAILNRQGLRVAVIVN
DSEVNLDAESVQDVSLHRGDELTEMSNGICCTLRADLEQISLARQORFDYLL
IESTGISPEMPVATFAFLDTREGSLRLADLTIVTVYVQSGFOALLSTDTVARAD
TEAHTSTRHLADLLIEQVYANVLVNRKDLIDEPGYQAVHAILAGLNPASIRMPMAH
GNVALSLDTHLFDLPASAPGWRKMEATDIPASSDTIGVTSWVTRKAPRHPQ
RLLEFLQKPNHGRLLRSKGYFWLASRHLIEIGLAAQSGKQFQWVGRWVNFIBSQW
PRDEFLOGIMAKWSVVGDCQELVFTGGQLDTRVLQRELDHCLLSAQEIAAGPLAW
QALPAATAPDTEALGARPTPPNAVQPT"
4439. .4443
/notes="SD sequence for OrfE"
4450. .5215
/notes="unnamed protein product; OrfE"
/codon_start=1
/transl_table=11
/protein_id="BAA14248.1"
/db_xref="GI:2160275"
/translation="MLFSTRTRWVLAPLADRLDOPDOACTPLSSNNPLRLRLTVE
QLLOEBSALKEQABALTOEVTOLDREVCARDLLEQWEARWALVSHGAGELFWELEVN
GTAAPSILACAMWTGASILAGIDQLGHNELHPADQRHLDIARHLADRSRGT
FVLVDRIQPPSGDDYRWCRISGDSRRDAQGLPVMGGSURDIHQHQLQDEVLELAIR
FDISREMLHDGLWDIEVWAGDPANPKNTIWWSSQMR"

```

ORIGIN

```

Query Match 14.2%; Score 224.8; DB 1; Length 5215;
Best Local Similarity 48.8%; Pred. No. 3.9e-20;
Matches 731; Conservative 0; Mismatches 752; Indels 15; Gaps 4;

Qy 37 CCACAGCCGACACAGCTCCAGGAGTACAGCGCCGCCACCTTCACCTTCAGCAGAGAA 96
Db 269 CCTACCTCCACACAGCTTTTAGACATCCGAAACCCAGCTTCACATGCACTGACGCGAA 328
Qy 97 CTGCGCCGCCAGCTCCGTCCTCGGTCGCGGAGATGGTACTGCTTCGACCTGATCGAC 156
Db 329 CAGCAGCGTCTTACCTGGAATGATGCAACGAGTTTCGACGCCCTACGACCTGTCGAC 388
Qy 157 GAACACTCCGCAACCGCGCAGCGCGCGCGCTACACGGAACCGGACATCGGCCCGCGAA 216
Db 389 GAACTGCTGATTTCGTTCCGCCAATACGCTACGACCGCAGTTCAGGCTATCGCCATCG 448
Qy 217 CCGACCGCGCACGACACCGGTTTCAACGATTCATCGGTTCTGCGGGTGGAGGCGCC 276
Db 449 CCATCGGCCCAAGGAAACCCCTCTGAACGCGCTGTGTACTACCGAACAGAGTGAATGTC 508

```



```

/ gene="VNG0852C"
993. .1643
/ gene="VNG0852C"
/ note="conserved hypothetical protein"
/ codon_start=1
/ trans1_table=11
/ product="Vng0852c"
/ protein_id="AAG19300.1"
/ db_xref="GI:10580420"
/ translation="MSVIAELSMSPRMILYEATTSVPNTFDVESVDVADGGDTVK
VAVAGDLSADDDADSTTTSVTLDDQADRELXSYQNADASEVMYEMLDLGA
AQLHVESQNGWVFVRPDRDALASFEICTSQVATLHCLISGVPEDATHQDPL
TDAQREAVRALDVGLVDPFRSASLSAVSEQLVGTQESERLRATRNLAANVD"
complement(1652. .2050)
/ gene="VNG0853C"
complement(1652. .2050)
/ gene="VNG0853C"
/ note="conserved hypothetical protein"
/ codon_start=1
/ trans1_table=11
/ product="Vng0853c"
/ protein_id="AAG19301.1"
/ db_xref="GI:10580421"
/ translation="MSVETDAAAGPSADHAADLARQLGDAIADLPYRRRLMETKEAVE
T'DEVQSIIEPEQRLQRFMLARQSGDADALNRVQEAQQHLHDHPVMADYLDQAQA
LEARFEGNLNLLSDLDVDFVGESGACQD"
2158. .3192
/ gene="VNG0854C"
2158. .3192
/ gene="VNG0854C"
/ note="conserved hypothetical protein"
/ codon_start=1
/ trans1_table=11
/ product="Vng0854c"
/ protein_id="AAG19302.1"
/ db_xref="GI:10580422"
/ translation="MSNEEARSAGDLNVTGSLRHRTDWTYELDRIVDEVERKDAEYV
GQFPEGKGFVADDLRLDTRIMLSQPCYGCACDLDTYLMRRTDVFVHFHGS
PKMESDIIYPLFQNVDPVEPTMOEARDELHDGVLVTTAQMNOPEMETVLEERG
YVTHRRGDRLTNEGVLGCNYSADIDAOVLVGGGKPEPLGLAMEHEPKNVLIA
DPVNAVSTADAEQILKERYASVHKMDAEOGVLPCTKIGQGRFDQAEELVENNDNA
YLITWDEVTPELDMFDMDAFVNTGCPRIITDDGQFQHKPMLTFOEYELAVGNPLED
LSFTFHGTW"
3592. .3861
/ gene="VNG0857C"
3592. .3861
/ gene="VNG0857C"
/ note="conserved hypothetical protein"
/ codon_start=1
/ trans1_table=11
/ product="Vng0857c"
/ protein_id="AAG19303.1"
/ db_xref="GI:10580423"
/ translation="WNPPFGAODGAHADRAFLAAASVAAVSXSVHNAGSKAFVESP
AGDRGTVTHAPAEFAVPAQSFHTQDRAVLDELFRVDWGRQS"
complement(3852. .5648)
/ gene="VNG0858C"
complement(3852. .5648)
/ gene="VNG0858C"
/ note="conserved hypothetical protein"
/ codon_start=1
/ trans1_table=11
/ product="Vng0858c"
/ protein_id="AAG19304.1"
/ db_xref="GI:10580424"
/ translation="MHVPSDAAAVALTAVGGVLAVSVVPPFRFARMLAGAAVSGVLA
GLVAGSSGLLDMRAVGAAYVAVLLATGTAFTVRGPQDQAVQAVRSLLTFGPWGLT
LVAGIVAGFYTLVQRPDGTLLVFPFISWSYAYPLGVLTPIAHANIGHVTLNIGTL
ALAPVAYAFSHVPTARGTAAGSRSTNPYPAGVVPFAGVLAVALLTSVPSMAVNG
FSQWVPAFGALLKYPLATVVAARDAISVFWETLLEPVPFASASQSPFPWAGV
AVQGHLLFGLLGAALAVLVHRDDHPARVWLGAVVIATSLSLWAVVWYGRGAEYV
LFRAGLLLVVLAVVAVAGTHGGVTLFPGRGVGTATARKLAAALLIIPVWMAFV
AVPVNLTTCQDANVAQSAVSNGYEVTAENATNKRVAADVIPYSGAATNVSASGVIV
/ gene="VNG0860G"
/ note="synonym: VNG0860G"
5748. .6032
/ gene="rpoL"
5748. .6032
/ gene="rpoL"
/ notes="rpoL"
/ codon_start=1
/ trans1_table=11
/ product="DNA-directed RNA polymerase subunit L"
/ protein_id="AAG19305.1"
/ db_xref="GI:10580425"
/ translation="MDLRIEIKDDAELTIEVAGENHTFMNLKGTILLETDGVTAASYD
MNPEQSGQTEPLLTIKTDGDDPVDALQAAAAGHT"SKLQSFADTPAAV"
complement(6039. .6860)
/ gene="hisP"
/ note="synonym: VNG0862G"
complement(6039. .6860)
/ gene="hisP"
/ notes="hisP"
/ codon_start=1
/ trans1_table=11
/ product="imidazoleglycerol-phosphate synthase"
/ protein_id="AAG19306.1"
/ db_xref="GI:10580426"
/ translation="MTLTKRVPICIDVDLDDGEPAYVTGVNPFELAYTGDPVEMAKR
YNEAGADEFVLDITASAEGRETMLDTVSADAVEFVPLTVGGGIRDTDIRTLRAG
ADKVISNGAIDAPSLVDRGAKAFSGQCIVISVDARRRDEQGHYTVQDGBSCWPEC
TVHGGREGTMDAIEMWQEAQRGAGELFVNSIDADGTQDGYDVPVLTAAVCDVASTPV
IASSGCGAPQDMADAYDAGADAALAASIHFHDEYSIAETKETLADAGYPIRAPDA"
7104. .7403
/ gene="VNG0861H"
7104. .7403
/ gene="VNG0861H"
/ notes="hypothetical protein"
/ codon_start=1
/ trans1_table=11
/ product="Vng0861h"
/ protein_id="AAG19307.1"
/ db_xref="GI:10580427"
/ translation="MARDRWLGADSFVGTVAAIAAADVHIEIAPGVPEVEVLDA
APRGVSRVTRVRELAVDAGATVVFVGSNDAGHVVGTLTVSGQGVASDDRDVHIVR"
7456. .7593
/ gene="VNG0863H"
7456. .7593
/ gene="VNG0863H"
/ note="hypothetical protein"
/ codon_start=1
/ trans1_table=11
/ product="Vng0863h"
/ protein_id="AAG19308.1"
/ db_xref="GI:10580428"
/ translation="MSDDHGDGGRVTAPMQBFSGVQTNGLIVLLVGLAVAYAVPVV
V"
complement(7601. .9703)
/ gene="purL"
/ note="synonym: VNG0864G"
complement(7601. .9703)
/ gene="purL"
/ notes="purL"
/ codon_start=1
/ trans1_table=11
/ product="phosphoribosylformylglycinamide synthase I"
/ protein_id="AAG19309.1"
/ db_xref="GI:10580429"
/ translation="MSLGSDDRVRVVAALGRDTPAAALFENLWSEHCAYRSRSRLI
Query Match 13.88; Score 218.2; DB 1; Length 15355;
Best Local Similarity 49.88; Pred. No. 2e-19;
Matches 741; Conservative 0; Mismatches 713; Indels 35; Gaps 6;
```

QY 40 A G A G C G A A C G G T C C A G A G T A C A G C G C C C G C C A C C A C T T C G A C T T C G A C G A G A A C T G 99
Db 13334 A C G T G T C G A C A C T G T G G G A C A T C C T C G A A C A C T T C G A G C G C T C G A A G A G T C C T 13393
QY 100 G C G C C C A G T C G T T C C G G T C G T G G G G A G A T G T G A C T C C T T C G A C C T G A T C A C A G A 159
Db 13394 G A A G T C G A C C G A G C C G A G T G T G A C G T G A T C G C G G C C G A G A G A T C C G A C T C C 13453
QY 160 C T A C C G A A C C G C C G A G C C G C G A C C G C G T A C G G A C C G G A C A T C G G C C G A A C C G 219
Db 13454 C T C G A C A G A C G A C G C G T C G A A A C G C C C G A A C C G A A C C G A A G A G C C G G T T C A A A G G C 13513
QY 220 A C C G G G -- A C G A A G A C C C G T T C A A C G C A T T C A T C C G G T T C T C C G G G T G G A G G C G C A 277
Db 13514 C C C A A G T C T G T A T G A C C T G A A C G C A T T A T A C C C A G A G A C C A T C C A G A G A G C G 13573
QY 278 C G A G G G C A C T A G A C G A C T T G A C C G C C G A T C A A G A C T G C A T C G C C A T C G C G G T A 337
Db 13574 C G A C G C C C C T C G A C G G C A C A C A C G T G C C G T C A A G A C A C A T C T C C A G A A G G C G 13633
QY 338 T G C C C A C C A G A C G G T C C G A T C C C G A C T C C C G A C T G A T C C C A C C G A G G A T C C G T G 397
Db 13634 T G C C C A C C A G T G G C T C A A G A T C T G G C C G A C T A C A G C C G C G T A C A C G C C A C C G 13693
QY 398 T G T G A G C G G T G T C T G C G G C A G G C C C A C C A T C G T C G G A G A C G A A C C T C G A G A C A 457
Db 13694 T C G T C A G A C C T C C T C G G G C C G C G G A C C A T C G T C G G C A A G A C C A C A C A T G G A C G A T 13753
QY 458 T G C G A T G G T A T C G T G A G G C A G C G T C T A --- C G G T C T C G C T G A A C C C G A C A A C C 514
Db 13754 T C G G C A T G G G A C G A C A C G A C C T C G T A C T T C G G T C C C A G A A C C C G G T C G A C G 13813
QY 515 C C G C C C A C G C A C G G T G G A T C T T C A G C G G G T C C C G C G T G C G C G T G C G C G C G A T G 574
Db 13814 A A C C C G G T G C C G G G C T C C T C G G C G G T C C G C G C C G T C G G A C G A C G C G A C G 13873
QY 575 T C A C T T C G C C T G G G G T C G A T A G A G A G A C A T C C G A T C C C G G C G A T G T G C G 634
Db 13874 C C A C C T C G G C T C G G T C C G A C A C C G G C G T C G G T T C G C G C C C G C G C G T A C T G T G 13933
QY 635 G A C T G T C G C A T G A A G G C A C C A C G C C T G T G C C T T A C G G C C T G A C A T A C A T G C 694
Db 13934 G C G T C G C C T C A A C C A C G A C T A C G G C T G T C T C C G G T A C G A C T G T G G C G T A C G 13993
QY 695 A C A C A C T T G A C C A C A T C G G C C C A T C A C A G G G G T C G A G T C A A C G C C C G G T C C 754
Db 13994 C A A C A C C T C G A A C A G A T C G G C C C A T C G C C C A C C C G T C G A G G C G G C G G A C T C C 14053
QY 755 T C G A G T G T T G C G G C G A C T G G C G A C C C T C A G T G G T C G T A C C T T C G G A C 814
Db 14054 T G A C G T A T C G C G C C C C A C A C A C A C A C G C A C C C C G A C G C G G C G G C G A G C G 14113
QY 815 C G A G A A C T A C G C T C C G C T C G G C A G G A G T A T C C G T C T G A G A T T C G C G T C G T C G 874
Db 14114 C T G A C T A C G C A C G C C G A C C G G C A --- C G T C G A T G G T T G A C C A T C G G C G 14164
QY 875 A G A G T C A C T G G A G C C G A C G T C G A C C G C G A C G T A T C G C G G T T C A C C A G G A C 934
Db 14165 T C C G C A G A A C T A C C A C G A C G C G T G A G A C C G C G T C G A G C C T T C G A A G C G C G 14224
QY 935 T G C G C G C T C G A G A C C G C G T C C A C C A C T C A G C G G T C T C G T C C G T T G T G A C G G 994
Db 14225 T T C G G A C C T C C G C G A G G C G C G A C A C C G T C G A C G T G G G T T C C G T C T G T G A T 14284
QY 995 C G C C T G C C A T C A G A G G G G T G A T G G C T T C A A C G C G C G C T T A G G G A C T C C G 1054
Db 14285 A C C C G T C G C C G C T A C T A C G T C A T C G C C A T G T C C G A A G C C T C T C G A A C C T C G C C G T 14344
QY 1055 C C G T G T G G G T A C T C C A A A G G G G C G T G A C G T C A C A C C C G C G T C A C A C C G C G G 1114
Db 14345 T C G A C G C G T C C G T A C G G C A C C G C G C G C G T T C A C G C A C T G A A C G A C A C G T T C G 14404
QY 1115 C C A G A T C G A C A C C A C A A G A T C G C G A T C T G T C C C G G C T G A T G T C C G G T A T C G 1174

Db 14405 C G A C G C G C G C G A C G G G T T C G G G A C A G A G T C A G C C C G A T C C T C T G G T A --- 14461
QY 1175 C G A G A C C T C G C G A C G A A T A C C T C G C A T C C A C T A C G G A A G C G C A G A C C T C G C G C 1234
Db 14462 C G T A C G C G T C G C G G G C T A C C A C A A T A C T A C A A G A A G C C C A G A G C C C G G G 14521
QY 1235 T G A G C T C G G A A C A G A T C A G C C G T C C T C A G A C C G G T G A C T G C T G A C C C G A 1294
Db 14522 C G T G G T G A A C A G G A C T T C G A G A G A C G T T C G C C G A C G T C G A C G T G G T G G C G T C C C C G A 14581
QY 1295 C C A G C C T A C C G T T G C C A A C A G A C T T G A G C G T C G G C A A G C A C C A C C A T G T C C A T G A T C C 1354
Db 14582 C G A T C C G G T C C C C C C G A A G C T C G G C A G A G C C T C G A C A C C C C G T C C A G A T G A T C 14641
QY 1355 C A C G G A T A C G G G A A T G C G A T C C T C A A C A G T C C C G C T G G A C C T C A C G G T C A C C C G G 1414
Db 14642 ----- T C G C G A C C C A A C C A C C A C C G C C G G A A C C T C G C A A C C T C C C C G 14686
QY 1415 C G T G A C G G T G C C C A C C G G T C G G G C G A A G A G G C C T C C C G T T G G C T C C A G T A T A G 1474
Db 14687 C A T C T C G G T G C C C G C G A C G C G --- A C G C C T C C C G T C G G C G T T C A G C T T C G 14743
QY 1475 G C C C C A C T T C G A G A G T C A C G C T T A C C G A C C G C G C G T A T C G A 1523
Db 14744 C C C G A A G T T C G G C G A G A A A C G A T C A T C A A C C G C C G C G C G T C G A 14792

RESULT 8
SC0939124 308050 bp DNA linear BCT 11-FEB-2003
LOCUS Streptomyces coelicolor A3(2) complete genome; segment 21/29.
DEFINITION AL039124 AL031031 AL031035 AL031124 AL031225 AL031231
ACCESSION AL031260 AL034447 AL035559 AL035569 AL0355913 AL645882
VERSION AL939124.1 GI:2413886
KEYWORDS Streptomyces coelicolor A3(2)
SOURCE Streptomyces coelicolor A3(2)
ORGANISM Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales; Streptomycineae; Streptomycetaceae; Streptomyces.
REFERENCE 1 Bentley, S.D., Chater, K.F., Cerdano-Tarraga, A.M., Challis, G.L., Thomson, N.R., James, K.D., Harris, D.E., Quail, M.A., Kieser, H., Harper, D., Bateman, A., Brown, S., Chandra, G., Chen, C.W., Collins, M., Cronin, A., Fraser, A., Goble, A., Hidalgo, J., Hornsby, T., Howarth, S., Huang, C.H., Kieser, T., Larke, L., Murphy, L., Oliver, K., O'Neill, S., Rabinowitsch, E., Rajandream, M.A., Rutherford, K., Rutter, S., Seeger, K., Saunders, D., Sharp, S., Squares, R., Squares, S., Taylor, K., Warren, T., Wietzorrek, A., Woodward, J., Barrell, B.G., Parkhill, J. and Hopwood, D.A.
Complete genome sequence of the model actinomycete Streptomyces coelicolor A3(2)
JOURNAL Nature 417 (6885), 141-147 (2002)
MEDLINE 21996410
PUBMED 12000953
REFERENCE 2 (bases 1 to 308050)
Bentley, S.D.
Direct Submission
Submitted (09-MAY-2002) Submitted on behalf of the Streptomyces sequencing team, Sanger Institute, Wellcome Trust Genome Campus, Hinxton, Cambridge CB10 1SA E-mail: sdb@sanger.ac.uk
On or before Oct 26, 2002 This sequence version replaced gi:20520664, gi:20520751, gi:20520752, gi:20520665, gi:20520811, gi:20520756, gi:20520845, gi:20520760, gi:20520817, gi:20520818, gi:20520782.
FEATURES
source location/Qualifiers
1. 308050
/organism="Streptomyces coelicolor A3(2)"
/mol_type="genomic DNA"
/strain="A3(2)"
/db_xref="taxon:100226"
250..2490
/gene="SC05495"

gene

```

/feature="synonym: SC8D9.07"
250. .2490
/feature="SC05495"
/feature="SC8D9.07, possible membrane associated
phosphodiesterase, len: 746 aa; similar to many eg.
TR:087376 [EMBL:AF052518] c-di-GMP phosphodiesterase A
from Acetobacter xylinus cdg2 operon (752 aa) fasta
scores; opt: 359, z-score: 1059.2, E(): 0, (31.7% identity
in 603 aa overlap). Appears to have two distinct regions
with the first 300aa containing several potential membrane
spanning hydrophobic regions and the remainder of the
protein showing similarity to the phosphodiesterase
mentioned above. Also similar to a nearby gene on the same
cosmid (SC8D9.22) fasta scores; opt: 736, z-score: 635.6,
E(): 5.3e-30, (32.6% identity in 729 aa overlap). Contains
TTA leu codon, possible target for bida regulation. Also
contains Pfam match to entry PF00990 DUF9, Domain of
unknown function, score 166.70, E-value 3.9e-46 and Pfam
match to entry PF00563 DUF2, Domain of unknown function 2,
score 291.30, E-value 1.2e-83. Contains possible
hydrophobic membrane spanning regions"
/codon_start=1
/transl_table=11
/product="putative phosphodiesterase"
/protein_id="CAB37571.1"
/db_xref="GI:4467246"
/db_xref="GOA:Q9Z584"
/db_xref="SPTREMBL:Q9Z584"
/translation="MEPTSAAGSRLLRLRWADAHVSRWLGORDGVSSVTGTRV
PCAVRPHGHQCHGHGLVDPDGRHLSWPALPAVNTAATFVLGAGFYAFSDGHAL
VVVLGVIARRRGQVGLHVLMLGRSGWGTGSGAALTAVILLVYGNIPAGVSVUT
KVLVAVAYLAWITALLWYIQTFRPGLPTAARTALVRLHGLVALLGIAPLICVVA
KPLLPFLAIPLALDSTLTARAEAEQLRDLPLTGLPNRWMLRTWLTALDDAERIG
ARAAMLLDIFRFRSVNTLGLHAGRLDRLLOTADRLCALPRGAEEARLGGDEFVILL
PVADSTTSATRSARLALSPMDLGLTAVLEASGAVFPDHALDAEGLRRADV
RMVQAKRRTGVVEVSRDNTSDPRLGLLGLRLALDAHEVQLHQPVRPDCQVAG
LEALVRVHPKRVKPPDEFIAIAESSGMLPHTLEYVLETLAQVQVRAQGLFVFPV
LVNSPRDVTTFGFAAGAAARLARGVPAQALQLEITEHVLEDFQRAADTLNALTGH
VMSJDDGTGSSYVHLRLFPVSELKIDRSFVARLAINDEDABIVCTVDLAHSLGL
LVVAGVGGDDTETWERLRLDGLCDVQGLVLAAMPPEETAWLLARGSGRWRAAALP
RAASDE"
1141. .1635
/feature="SC05495"
/feature="Pfam match to entry PF00990 DUF9, Domain of unknown
function, score 166.70, E-value 3.9e-46"
1420. .1422
/feature="SC05495"
/feature="TTA leu codon, possible target for bida regulation"
1672. .2406
/feature="SC05495"
/feature="Pfam match to entry PF00563 DUF2, Domain of unknown
function 2, score 291.30, E-value 1.2e-83"
2774. .3145
/feature="SC05496"
/feature="SC8D9.08"
2774. .3145
/feature="SC05496"
/feature="SC8D9.08, small hydrophobic membrane protein, len:
123 aa; no significant database similarities. Contains
possible hydrophobic membrane spanning regions"
/codon_start=1
/transl_table=11
/product="putative small membrane hydrophobic protein"
/protein_id="CAB37572.1"
/db_xref="GI:4467247"
/db_xref="SPTREMBL:Q9Z583"
/translation="MTVTAYTLAVVLNLFCLFLGYRFLFPQASATGYGVPADAGGDA
RAYLTVKGVKRGDTGLGVGLALAVAGAPPEARFWMFLCVLVALPLADTLVLRHGGTKVA
FGIHFATAIIVLISAGLLFAV"
3177. .3671
/feature="SC05497"
/feature="SC8D9.09, unknown, len: 164aa; similar to many of
undefined function eg. TR:O06336 [EMBL:Z95390]
hydrophobic protein from Mycobacterium tuberculosis (177
aa) fasta scores; opt: 171, z-score: 213.4, E(): 0.00014,
(34.9% identity in 109 aa overlap)."
/codon_start=1
/transl_table=11
/product="conserved hypothetical protein"
/protein_id="CAB37573.1"
/db_xref="GI:4467248"
/db_xref="SPTREMBL:Q9Z582"
/translation="MSLVFKPDESVVVRREAEVVGRAPTTVRLADSSSTGGALST
QAVTUTAGADGAKPHWDSAMFPLLGADADVLGDEVLTAGPDGLIVPRGKPHAF
RAVVGADADLLIVLAPGHWVERFEYFRHLQRIALGEATPESLLEVEYQELYNHFLSSAAMD
ARRG"
3816. .4112
/feature="SC05498"
/feature="SC8D9.10, probable Glu-tRNA-Gln amidotransferase
subunit C, gatac, len: 98aa; similar to many eg. TR:O06492
[EMBL:AF008553] Glu-tRNAgln amidotransferase subunit C,
gatac, from Bacillus subtilis (101 aa) fasta scores; opt:
232, z-score: 294.8, E(): 4.1e-09, (42.1% identity in 95
aa overlap)."
/codon_start=1
/transl_table=11
/product="probable Glu-tRNAgln amidotransferase subunit C"
/protein_id="CAB37574.1"
/db_xref="GI:4467249"
/translation="MPGITREEVAAHLARLARLELKPPELEHFGALDDITIGAVARVSE
VADQVPPTSHPLPLTNWVRPDEVPSTLTPQALSGAPAEQQRFKVPQILGEE"
4116. .5609
/feature="SC05499"
/feature="SC8D9.11, probable Glu-tRNA-Gln amidotransferase
subunit A, gata, len: 497aa; similar to many eg. TR:O06491
[EMBL:AF008553] Glu-tRNAgln amidotransferase subunit A,
gata, from Bacillus subtilis (486 aa) fasta scores; opt:
1443, z-score: 1512.5, E(): 0, (52.0% identity in 465 aa
overlap)."
/codon_start=1
/transl_table=11
/product="probable Glu-tRNA Gln amidotransferase subunit".
/protein_id="CAB37575.1"
/db_xref="GI:4467250"
/translation="MSDIILKLTAAEIAAKIASGELTAVQVTEAHLARIEAVDEKTHAP
LHVDRGALAAQAVDEKKEKGLGLAGVPLAKDITFTTEGIPITTVGSKLEGWIP
PYDATVTKLKAADVILKTNDEFAMSGSTENAYGTPNPDLTKIPGSGGSS
AALAFQAPLAIGTDTGGSIQTPASVTGVTGVTGYGVRVYGVAFSSSLDQGGPCA
RTVLDALLHEVITAGHPMDSTSIDAPVAVVEAARNGSVDMRVGVKQFQEGYQA
GVVQREFDESVELLSIGARIVELDCPSDLALSAYVLIAPSSCSNLARFDCGLRYGAR
VGDGTHSAEVTSLTREAGFGEVKRIMLTGYALSISGYDAYVYSSAQKVTLLKQE
FERAFQGVIVTSPPTTPTFAIGRADDFMAWYLDLCTIETIEPLNAGNAMSPLCGLA
PEDNLPEVQIIPAPAMKDDRLYKVGAAVEAFVEKWHPLIEEAPSL"
5606. .5845
/feature="SC05500"
/feature="SC8D9.12"
5606. .5845
/feature="SC05500"
/feature="SC8D9.12, putative membrane protein, len: 79 aa.
Contains possible hydrophobic membrane spanning regions"
/codon_start=1
/transl_table=11
/product="putative membrane protein"
/protein_id="CAB37576.1"
/db_xref="GI:4467251"
/db_xref="SPTREMBL:Q9Z579"

```



```
/translation="MSALTAKGFKKSKGLVYSMATSAFAGVAVYKQIKKARGESDT
LRLLDAVVTGAAVVTNLAITYRELKELGDDVLLG"
5861.. 7375
/genes="SCO5501"
/notes="synonyms: gatB, SCO9.13"
5861.. 7375
/genes="SCO5501"
/notes="SCO9.13, probable Glu-tRNA-Gln amidotransferase
subunit B, gatB, len: 504aa; similar to many eg. TR.O30509
(ENBL.AF008553) Glu-tRNA-Gln amidotransferase subunit B,
gatB, from Bacillus subtilis (476 aa) fasta scores; opt:
Query Match 13.2%; Score 208.4; DB 1; Length 308050;
Best Local Similarity 49.8%; Pred. No. 1.6e-18;
Matches 651; Conservative 0; Mismatches 631; Indels 24; Gaps 4;
262 CGGGTGGAGGCGCCACGAGGCGGCACATGAGGACCTGACCGCGGATCAAGACTCG 321
4302 CGGAGCGCGGAGAGTCTGGTTCGCTGCGCGGCTCCCTCGCGCTCAAGGACATC 4361
322 ATGCCCATCCCGGTATGCCACACAGAAAGGGTCCCGGATGTCCTCCGACTGTGATCGCC 381
4362 TTCACACCGAGGACATCCACACACCGCTCGGCTCGAAGATCTCGAGGCTGGATCCG 4421
382 ACCGAGATCCGCTGTGTGGAGCGGCTCTCGCGGAGCGGCCACCATCTCGGCAAG 441
4422 CCGTACGACGACCGCTCAAGAGCGCTCAAGGCGCGGACGTCTGCTATCTCGGCAAG 4481
442 ACGAACCTCGAGGACATGGCGATGGG---TATCGGTGAAGGACAGCTCTACGGTCTCTCGG 498
4482 ACCAATGACGAGTTCGCCATGGGTCTCTCCACCGAGAACAGCGCTACGGCCCGAC 4541
499 CTGAACCGAACAACCCCGCCCGGCAAGGGTGGATCTTCCAGCGGCTCCGCGGTGCG 558
4542 GGCACACCCCTGGAGCTTCAACAGATCCCGCGGCGTCCGCGCGCGCTCTCCGCGGC 4601
559 GTCCCTCCGCGATGTCGACTTCCGCTCGCGGTGCGATGAGGACGACATCCCGGATC 618
4602 CTGCGCGCTTCAGGCGCCCGCTCGCATCGGACCGACCGCGGTTCATCCCGCAG 4661
619 CCGCGCGCATGTCGAGCTGTCGCGATGAAGCGGACCAACCGCGCTGTCGCTTAC 678
4662 CCGCGCTCGCTACCGGACGCGTCCGCTCAAGCGGACGATACGCGCGGCTCTCCGCTAC 4721
679 GGCTGACATACATGACACACTTGGACACATCGGCGCCATCACCGGGGGTTCGAG 738
4722 GGCATGTGGCTTCTCTCTCTCCGACGAGCGCGCCCTGCGCGCGGCTCTCCGCTG 4781
739 CTCAACCGCGGCTCTCGAGGTGTTGGCGGCGCGGCTGCGCGGACCTCGAGTGGTG 798
4782 GACCGCGCTCTGTCACGAGGTGATCGCGCGGCGACGACCGGATCCACCTCCAT- 4840
799 CGTAACTCTCGGAGCGGAGAACTACGGTTCGCGCTCGCGAGGAGTATCGGCTG 858
4841 --CGACGCGCCGCTCCCGCGGCTGTCGAGCGCGCGCGCAACGCGCTCGAGCGGATG 4898
859 AGATTCCGGTCTCGAGGAGTCACTGAGCGGACCGGACCGGCGGACGCTGATCGCC 918
4899 CGCTCGCGCTCTGTCAGGAGTTCGCGCGGAGGCTTACAGCGCGGTGTCGTCAGCGC 4958
919 CGGTTCAACAGGAGTTCGCGCGCTCGAGAGCGCGGTCGACCATCGAGCGGCTGTCG 978
4959 TTCAGCAGTCCGTCGAGTCTGTAAGTCGCTGGCGCGCGGAGATCGTCGAGCTGGATGC 5018
979 GTGCGGTTGTGGACGCGCGCTGCGCTTATCCAGAGCGGCTGATGGCTTTCAACGCGCGC 1038
5019 CCGTCTTCGACCTCGGCTCTCCGCTACTACCTGATCGCCCGCTCGAGTCTCTCC 5078
1039 GCTATGGCGGACTCCGCGGCTGCTGCTTCTTCCAAAGGCGCGCTGAGAGTCAAGCAC 1098
5079 AACCTCGCGGCTTCGAGCGCTCGGCTTACGCGCGCGGCTGCGCGAGCAGCGCACGAC 5138
1099 GCGGTACGACGCGGCGCGGCTGCGAGTTCGACACACCGACCAAGGATCTGCGCGG 1158
```

```
Db 5139 TCCGCGGAGGAGTCACTCCTCGTACCGCGGAGCGCGCTTCGCGCGGAGTCAAGCGC 5198
Qy 1159 CTGATGCTGTGATCGGAGACACTCTGGCGAAGAAATACCTCGCATCCATCGCGAG 1218
Db 5199 CGCATCATGCTCGGACGATCGCCCTGTCGAGCGGCTACTACGACGCGTACTACGGCAGC 5258
Qy 1219 GCGCAGAACCTTCGCGCTGGAGCTCGGCAAGCAGATCGACGCGCTCTCCAGGACGGGCT 1278
Db 5259 GCCCAGAGGTCGCGACCCCTCATCAAGCAGGAGTTCGAGCGGCTTCGAGCAGGTGAC 5318
Qy 1279 GCATGCTGACCGCGACCGCTACCGTTCGCGTTCGCGTTCGAGGCTGAGGCTGCGCAAGAC 1338
Db 5319 GTGATGCTCTCCGCGACCGCCGACCGCTTCGCGATCGCGGAGCGCGCGGAC 5378
Qy 1339 ACCATGCTCATGATCCCAAGGATGACGCGGCAATCGGATCTCTCAACAGTTCGCGCTGAC 1398
Db 5379 CGATGGGATGATCT-----GGCGGACCTGTGCACCATCCCAACAC 5423
Qy 1399 CTCACCGGTACCCGCGGCTGACGCTGCCACAGG---TGGCGGCGAGAGGCGCTGCGCC 1455
Db 5424 CTGGCGGCAACGCGGCGCATGCTGCTGCGGCTTCGCGCGGAGGACCACTGCGC 5483
Qy 1456 GTTGGCTTCAAGTATAGGCGGCGCATTCGAGGAGTTCGAGCTTACCGCACCGCGGCC 1515
Db 5484 GTGGTCTGCGATCATCGCCCGCATGAAGACGACCGGCTCTCAAGGTTCGCGCC 5543
Qy 1516 GTGATCGAGGCGCGGCTATGGAGCTTCGCGCGGAGCGCGC 1561
Db 5544 GCGGTCGAGGCGCGCTTCTGTAAGTGGGCGCACCGCTGATCG 5589

RESULT 9
AP005032/c
LOCUS
DEFINITION Streptomyces avermitilis genomic DNA, complete genome, section 12/30.
ACCESSION AP005032 BA000030
VERSION AP005032.1 GI:29606342
KEYWORDS
SOURCE
ORGANISM Streptomyces avermitilis MA-4680
Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales; Streptomycineae; Streptomycetaceae; Streptomyces.
REFERENCE
AUTHORS Omura, S., Ikeda, H., Ishikawa, J., Hanamoto, A., Takahashi, C., Shinose, M., Takahashi, Y., Horikawa, H., Nakazawa, H., Oonoe, T., Kikuchi, H., Shiba, T., Sakaki, Y. and Hattori, M.
TITLE Genome sequence of an industrial microorganism Streptomyces avermitilis: deducing the ability of producing secondary metabolites
JOURNAL Proc. Natl. Acad. Sci. U.S.A. 98 (21), 12215-12220 (2001)
MEDLINE 21477403
PUBMED 11572948
REFERENCE
AUTHORS Ikeda, H., Ishikawa, J., Hanamoto, A., Shinose, M., Kikuchi, H., Shiba, T., Sakaki, Y., Hattori, M. and Omura, S.
TITLE Complete genome sequence and comparative analysis of the industrial microorganism Streptomyces avermitilis
JOURNAL Nat. Biotechnol. 21 (5), 526-531 (2003)
MEDLINE 22608306
PUBMED 12692562
REFERENCE
AUTHORS Omura, S., Ikeda, H., Ishikawa, J., Hanamoto, A., Takahashi, C., Shinose, M., Takahashi, Y., Horikawa, H., Nakazawa, H., Oonoe, T., Kikuchi, H., Shiba, T., Sakaki, Y., Hattori, M., Kikuchi, H., Shiba, T., Sakaki, Y. and Hattori, M.
TITLE Direct Submission
JOURNAL Submitted (29-MAR-2002) Director-General of Biotechnology Center, National Institute of Technology and Evaluation, Biotechnology Center, 2-49-10 Nishihara, Shibuya-ku, Tokyo 151-0066, Japan (E-mail: bio@nitech.go.jp, URL: http://www.bio.nitech.go.jp/, Tel: 81-3-3481-1933, Fax: 81-3-3481-8424)
```

COMMENT

This work was done in collaboration with Haruo Ikeda(*1), Jun Shikawa(*2), Akiharu Hanamoto(*3), Chikazu Takahashi(*3), Momomi Shinose(*3), Hiroshi Horikawa(*4), Hidekazu Nakazawa(*4), Tamomi Osonoe(*4), Norihiro Kushida(*4), Hisashi Kikuchi(*4), Tadayoshi Shiba(*5), Yoshiyuki Sakaki(*6,*7), Masahira Hattori(*1,*7) and Satoshi Omura(*1,*3).

Final finishing process and all annotation were done by H. Ikeda and J. Ishikawa.

- *1 Kitasato Institute for Life Sciences, Kitasato University
- *2 National Institute of Infectious Diseases
- *3 The Kitasato Institute
- *4 National Institute of Technology and Evaluation
- *5 School of Science, Kitasato University
- *6 Institute of Medical Science, University of Tokyo
- *7 RIKEN, Genomic Sciences Center

Following url is also available.
<http://avermiltilis.ls.kitasato-u.ac.jp>.

FEATURES

source

1. 281450

/organism="Streptomyces avermitilis MA-4680"
 /mol_type="genomic DNA"
 /strain="MA-4680"
 /db_xref="taxon:227882"
 /note="This strain is also named as strain: ATCC 31267,
 NCIMB 12804 or NRRL 8165.-synonym: Streptomyces
 avermectinius"

gene

359. .4195

/gene="cvaA6"

359. .4195

/gene="cvaA6"

/note="SAV2694"

multi-component regulatory system-6"

/codon_start=1

/transl_table=11

/product="putative sensor-like histidine kinase"

/protein_id="BAC70405.1"

/db_xref="GI:23606343"

/translation="MQGRFKRDSGSAEPEPHGGTDRGSSPQHAQNPQPGAGSGGR
 TGAPAKPPTSSGSSRVVALRNWTRISLVLALPVVAATSLGALRISDSM
 DEIQLDNMKLTMTKQATLAAALQDERQDSAGLAHGATATDFVKGVRKTDRA
 RSAPLSGTQEDDASGNSQSVRSVVLGAGLSNLSIRNAYKSENATQVEGY
 HKLTQLGLSDMAEATSNFEMIORTLAFASTAKESVQRAVIAASLPASNTP
 GKLSDNQYALSUDSEISFNSIYGDEYDITKPIDSGNPTIEADEYAKRAL
 SDNRIQONKRSYKDWYDDSDAKLEQMSKIELTLNQMEQKARELRNQSQAIIIG
 ALILIVGLVGAFTVARSIMRSLRIQDTATKVAQDLPELVQLSESPPQVDVTS
 VESVGHKDETGQVAAAFDDVHREAVLAAELLRGNVAMFTNLSRRSQGLIQRV
 LSLISELSREADPDLSSLFKLHLATMRNENGLAVLAGEPGRWTPRPVLVDV
 LPAASEVEQERIELISVPTTEVAGRVVNDLVHLLAELENHNSFSPQTKVKTGH
 ALDGRVIEIHDTIGLSPEDLAAINERLAAPPTVDVSVSRKMLFVVGRLSORHG
 RIQLRPSGGTTLVLPVVDVVAQGGKVPQGGAGFQGGPAAQAASAGAAARGI
 GGGSPSLGAGAPAGNAGRIGAAPQGGVQAGQGPRAALPGSNQGGPAGGARGPQ
 QGPAAPQGRPAPAGAGAGAGAFQGAQFQGLQAQAFNEGFAATQGGFGDGGSG
 TRGGGFGDTPAAAPQPOQPOQPOQPKESRRRRPOLPGRGPGRAELPGGSAQRPTRP
 WSDENAQRPVPSLDADEGHEEPDALSATSLRPVDRDQGPATSEIPRIDDRGPG
 ATSEAPDFDTTRCAEAPQSGNGQFGRPMFGNGQAAPNRPFAQRPQDASPTG
 QPPAVGGSDPQATQFATPGVDGSGTQYATPGDAGTQYFAPQGNQDQSSSTG
 QERPAQGFADFGVRRPAPAPQPPAPQPPAPPAQPPAPGPGDGTPLYLITLNTWPHQ
 CQQQQGSLGNSQAAPQPOQSPAPAPQPPAPPAASWRSTFNDLVQAEARQPSA
 GGVVTSGLPRPRVLANLVAGTAQQQHQAGPQVSRAPDDVRGLTNLRGIIQQGRQAG
 TQGTGSPSPTRQER"

gene

4205. .4618

/gene="cvaB6"

4205. .4618

/gene="cvaB6"

/note="SAV2695"

multi-component regulatory system-6, containing

roadblock/LC7 domain"

/codon_start=1

/transl_table=11

/product="hypothetical protein"

/protein_id="BAC70406.1"

/db_xref="GI:29606344"

/translation="MSQAAQNLMWLTITNFVDNTPGVSHTVVVSADGLLLAMSEGPFRD

RADQLAAVASGLTSLTAGASRIFFEGGTVAQTVMEMERGFLFLMSVSDGSLAVLSHPHE
 CDIGLVGEMALLVDRAGAVLTPDLRAELQGSLLH"
 4730. .5128
 /gene="cvaC6"

4730. .5128
 /gene="cvaC6"
 /note="SAV2696"
 multi-component regulatory system-6"
 /codon_start=1
 /transl_table=11
 /product="hypothetical protein"
 /protein_id="BAC70407.1"
 /db_xref="GI:29606345"
 /translation="MTPFPASHDPIVAQSYGDEGDQPLVRPVAVTGRTTRPRVQLAIEA
 LISTTADPAQLMGLLPEHQRIHCLREKVSVAEVSALLSMPGLVARILVADLAAGLV
 AIHQPGGDNNGGAPDVTLLERVLGSLRL"

gene

5109. .5687

/gene="cvaD6"

5109. .5687

/gene="cvaD6"

/note="SAV2697"

multi-component regulatory system-6"

/codon_start=1

/transl_table=11

/product="putative ATP/GTP-binding protein"

/protein_id="BAC70408.1"

/db_xref="GI:29606346"

/translation="MDPSSDGGRTTTSAKIIVAGGFGVGTGTFVGAISEINPLETE
 AVNTSASAGIDLTHTGDKTTTVMDFGRITLDQDLILFLGTQDQRFNFMWDLV
 RGAIGAVLVDTRRLADCPFVDPYFENSGLPFVIALNGDFGHPQYNPEEVRALQIGP
 DAPIITTDARHADRAKSAKITLVEHALMARLR"

gene

6085. .9330

/gene="cvaA7"

6085. .9330

/gene="cvaA7"

/note="SAV2698"

multi-component regulatory system-7"

/codon_start=1

/transl_table=11

/product="putative sensor-like histidine kinase"

/protein_id="BAC70409.1"

/db_xref="GI:29606347"

/translation="MRRSRNPEPSARGNFTPPRGCVPAQSGPEPAAVVAPSGSRLS
 PRNVRVTLNALILLPVVLVVGFGFQVKGSIIDTWQEADEAKVAKIVASAGEYABA
 LNMERVSAPALLKGRNNATVKKAFALTDQADAFPHQVGVMPKAGLORLGLVEK
 AEPQLQIKAAVTTAMDPKTEBEGYTVTHLIMEFSENLGLTGNTISYGRVYVIA
 LAGKASLRSISGLHLLIKPSSDEKIRAQOITSFTSYALEQIALQEVYSGGTQDYA
 RLTVMKQKTAAGKALAKAKSRTAAGTVPAPALDKMIAIGSGATPAALAOQGI
 TPTWMSAATLKFQSYQVEKELVDVANDANDIASSAQRDAYINGAIVVALLAIFI
 LAGLMARQMSRQRNNAFVNAEORLEMLVDQLSRDTPGRVDTRVQAIPTITDEI
 GEVARAPDVHREAVLAAEQALLRGNINAIPTNLSRRNQSLEGGQTLITLLENBA
 DPQLENLFDHLATMRNENGLVLAGEEFGRRWDQFVLVDVLRASSEVEQYI
 RVLSGVPAEAIHGRAVTDLVHLLAELENATTFSSPQTKRVTRATRLPDGRIMVRIH
 DKIGLTAEDPADINHLANPPTVDAAIISQRMGLFVVGRLSDRHGIRVLRPSGEQAG
 TTSVLMPDVIITHGGGQGVORDEFTVSQIIPQONFAGGGEAFQVPMWTAELGFI
 DSRYTEVDDTIELDPVGRSLNERRAALQAHPQDQSGEGYQDEFEASQCPA
 YGNGYEQEQAGYQQQYATHEQQQPAIEEYFAENGFSQGTGEFAPGAPNGDF
 ASNGVPEPSIAEPNSESNGGRTASFTSFAPDEFSQDQDMFKQDQISGYKSEVAP
 EAESSQAADYSERDEHYGTPGPAPSAHALTDAGLPRRGSTNGNGNGQGRVPPPS
 AAPAGTETGMSNRWSSNDRMQQASQKPKAGGTVSSGLPRRYPKANLVETGADS
 TPOGGPQVSAPEDVAGRLSNLRGVRGRENAGSTNGQGLQDPDSTYNQER"

gene

9341. .9754

/gene="cvaB7"

9341. .9754

/gene="cvaB7"

/note="SAV2699"

multi-component regulatory system-7, containing

roadblock/LC7 domain"

/codon_start=1

/transl_table=11

/product="hypothetical protein"

/protein_id="BAC70410.1"

/db_xref="GI:29606348"

```
/translation="MSQAAQNINLWLTINFDVNTPGVSHTVVVSADGLLIAMSEGPFRD
RADQLAAVASGTLTAGASRIPEGGAVNVTVMERGFLEFLMSISDGLSLAVLHPE
ADILGLVGYEMALLVDRAGTGLTPDLRAELQGSILN"
9851..10435
/genes="cvmC7"
CDS
9851..10435
/genes="cvmC7"
/notes="SAV2700
multi-component regulatory system-7"
/codon_start=1
/translation="MATPPGGSSGNWSYGPQGQGGQBNRYNFPSPRPQYPAP
13.2%; Score 208; DB 1; Length 281450;
Query Match
Best Local Similarity 50.1%; Pred. No. 1.8e-18;
Matches 641; Conservative 0; Mismatches 615; Indels 24; Gaps 4;
262 CGGTTGAGGGCGCCACGAGGGCCACTGAGCGACTGACCGCGCGATCAAGGACTGC 321
Db CGGAGCGCGCGAGAGCTCGGCCCTCGCGCGCTCGCGCTCGCGCTCAAGGACATC 72080
322 ATGCCATCGCGGTATGCCACACGAAAGGTCCTCGGATGCTCCGACTGTGATCGCC 381
Db TTTACACAGGAGGATCCCGACACCGCTCGGTTCGAAGATCTCGAGGGCTGATCCCG 72020
382 ACCGAGGATCGGTGTGTGGAGCGGCTGCTCGCGGAGCGCCACCACTGTCGCGCAAG 441
Db CCCTACGAGCGGACCTTACCAGAAAGCTGAAGCGCCGCGACTGCTGCTCGGCAAG 71960
442 ACGAACCTCGAGGACATGGCGATGG---TATCGGTGAAGGACGCTACGCTCTCTGG 498
Db ACCAACATGAGGAGTTCGCGATGGGTCTTCCACGAGAACAGCGGCTACGCGCCAC 71900
499 CTGAACCGGAACACCCCGCCACGACGCGGTGGATCTTCCAGCGGCTCCGGCGTGC 558
Db GGCAACCCCTGGGACCTCAGCGTATCCCGGTGGCTCCGCGGTGGCTCTCGCGCGC 71840
559 GTCGTCGGGATGTGACTTGGCTTGGCGTGGGCTGATGAGGAGGACGACNTCCGATC 618
Db CTCGCTCTTCGAGGGCCCGCTCGCATCGGACCGACACCGCGCGCTCCATCCGCG 71780
619 CCGCGCGCATGTCGGACTGTCGATGAAGCGACCCACGCGCTGTCGCGCTTCTAC 678
Db CCGCGCGCGTACCGGACGTCGGGTGAGCGGACGTCAGCGCGGTCTCCGCTAC 71720
679 GGCCTGACATACATGAGACCACTTGGACACATCGGGGCCATCACAGGGGGTGGAG 738
Db GGCTATGTTGGGTTCTCGTCTCCCTCGACAGGGCGGCGCCCTGTGCGCGTACGTC 71660
739 CTCGAACCGCGGCTCTCGAGGTGTGGCGGGGCGGACCTGGCGGACCTCGAGTGGTG 798
Db GACGCGGCTCTGTCAGAGGTATCGCGGGGACACACCGCTCGACTGAGTCCAT- 71601
799 CGTAACCTTCGAGGACCGGAGAACTACGGTTCGCGTTCGCGGAGGAGTATCGGTCG 858
Db -CGACGCGCGGTCGCGCGGTGTCGAGCGCGCGCGACGACGCGTCAAGGGCATG 71543
859 AGATTTCGGTTCGAGGAGTCACTGAGCGGACGTCGAGCGGACGTCGAGCGGATCGCC 918
Db CGGTCGCGGTCGTCAGAGTTCGCGGCGGAGGGCTACACAGGCGGTGTGCTCCAGCG 71483
919 GCGTTTCAACAGGAGTGGCGGCTCGAGAGCGCGGTCGACCATCGAGCGGCTCG 978
Db TTCGACGAGTCCGTCGCGCTCTCAAGAGGTGGCGCGGAGATCGTCGAGTGGACTGC 71423
979 GTGCGGTTCGAGCGCGGCTGCGCTATCCAGAGCGGCGTGTGCTTTCAACGCGCGC 1038
Db CCGTCTTCGACCTCGCGCTGTCGCGGCTGCTGAGTTCGCGCGCTCGAGTGTCTCTCC 71363
1039 GCTATGCGGACTCGCGCGGTGTGGGTACTTCCACAGAGGGCGCGTGGAGCTCAGCAC 1098
```

```
RESULT 10
LOCUS AE016864
DEFINITION Pseudomonas syringae pv. tomato str. DC3000 section 9 of 21 of the
complete genome
ACCESSION AE016864
VERSION AE016864.1 GI:28852711
KEYWORDS
SOURCE Pseudomonas syringae pv. tomato str. DC3000
ORGANISM Pseudomonas syringae pv. tomato str. DC3000
Bacteria; Proteobacteria; Gammaproteobacteria; Pseudomonadales;
Pseudomonadaceae; Pseudomonas.
REFERENCE 1 (bases 1 to 310325)
AUTHORS Buell, C.R., Joardar, V., Lindeberg, M., Selengut, J., Paulsen, I.T.,
Gwinn, M.L., Dodson, R.J., DeBoy, R.T., Durkin, A.S., Kolonay, J.F.,
Madupu, R., Daugherty, S., Brinkac, L., Beanan, M.J., Haft, D.H.,
Nelson, W.C., Daviden, T., Zafar, N., Zhou, L., Liu, J., Yuan, Q.,
Khouri, H., Fedorova, N., Tran, B., Russell, D., Berry, K.,
Utterback, T., Van Aken, S.E., Feldblyum, T.V., D'Ascenzo, M.,
Deng, W.L., Ramos, A.R., Alfano, J.R., Cartinhour, S., Chatterjee, A.K.,
Delaney, T.P., Lazarowitz, S.G., Martin, G.B., Schneider, D.J.,
Tang, X., Bender, C.L., White, O., Fraser, C.M. and Collier, A.
The complete genome sequence of the Arabidopsis and tomato pathogen
Pseudomonas syringae pv. tomato DC3000
Proc. Natl. Acad. Sci. U.S.A. 100 (18), 10181-10186 (2003)
JOURNAL 12928499
PUBMED 2 (bases 1 to 310325)
REFERENCE Buell, R., Joardar, V., Khouri, H., Fedorova, N., Tran, B., Russell, D.,
AUTHORS Berry, K., Utterback, T., Van Aken, S., Feldblyum, T., Gwinn, M.,
Dodson, R., DeBoy, R., Durkin, A., Kolonay, J., Madupu, R.,
Daugherty, S., Brinkac, L., Beanan, M., Haft, D., Selengut, J.,
Nelson, W., Daviden, T., White, O., Fraser, C. and Collier, A.
Direct Submission
TITLE
JOURNAL Submitted (03-MAR-2003) The Institute for Genomic Research, 9712
Medical Center Dr, Rockville, MD 20850, USA
FEATURES
Location/Qualifiers
source 1..310325
```


GVEVPARSGKHLRTVPAGERLITTAGELIKVESIKQIAOEFNSKEKGTLSIATHTT
QARYALPVIUNFNIQYEDVALHMQSPMQIAEMADGTVDFPALTAEALBQFSDLVW
MCIYRNKCVVVFQGRHUKVPLKLELLASIPVIVYVFGTGRSKDSEAFSHRGLVP
KVFTAADADVITKTIRUNLGVIVARWAVEEDVDLDVLDLSEFVSVKIAFR
GTFLEFMDIFIEKFAPLTRVWAKA:QCHNQEMSELFANVELPVH"
complement (7787. .8521)
/gene="cysh"
/locus_tag="PSPTO2280"
complement (7787. .8521)
/gene="cysh"
/locus_tag="PSPTO2280"
/codon_start=1
/transl_table=11
/product="phosphoadenosine phosphosulfate reductase"
/protein_id="AA055794.1"
/db_xref="GI:28852722"
/translation="MSQPPDVAEALATTYANKSADILKLAPEHFGDELWISFGAEDV
VLVDANKLNKNVPSLDTRGLRPETVRPIEQVDRDHYKIDIEIISPPORALEPFRKE
KGLF9YRDHGECGIRKIEPRRLKLSVKAATGQRRDQSPGTRSAVALEVD5AF

Query Match 13.0%; Score 205.8; DB 1; Length 310325;
Best Local Similarity 50.8%; Pred. No. 3.4e-18;
Matches 632; Conservative 0; Mismatches 592; Indels 21; Gaps 5;

QY 210 CGCGAACCGACGGCGGACGAGACCCGTTCAACGATTCATCCGTTCTCGCGGTGGA 269
Db 156648 CAGCAAGCCGGCAGATGCGCAAAACCTGCACGGTGCCTGGTACGTAAACCAACCATCGT 156707
QY 270 GGGCGCACCGAGGGGCCACTGAGCACCTGACCGCGCGGATCAAGGACTGCATCGCCAT 329
Db 156708 CGGTGCGCAACAGGTCCTCTGGCGGAGAACGGTGGTTCATCAGGCAACAGTTTCAGT 156767
QY 330 CGCGGTATGCGCACCAAGACGGTCCCGGATGTCCTCCGATGTGATCGCCACCGAGGA 389
Db 156768 GCGCGCGCTGCCGATGGCTAATGGTTCTTTGAGCCTTGATGGCTACGTGCCGCTCGAAGA 156827
QY 390 TGCCTGTGTGTGAGCGGCTGCTCGCGCAGCGGCCACCATCTCGTCCGCAACGAACT 449
Db 156828 CGCACGGTGTCAACAGCCTGTGCGCGCTGTGCGAGCGGTGTGCGCAATCATGTG 156887
QY 450 CGAGGACATGGCAT---GGGTATCGGTGAAGCAGCGTTCAGGTCCTGCGCTGAACCC 506
Db 156888 CGAAGACCTGTGTTCTCGGAGCAGCTTTACCTCGGCGAGCGCGCGGTGAACAAACC 156947
QY 507 GAACACCCCGCCACGCGACGGTGGATCTTCCAGCGCTCCGCGCTCGCTCGCTG 566
Db 156948 GTGGGACCTGGCCCGCAACCGCGCGCGCTCGTCCAGCGGAGTGGCTGCTGCTGCTT 157007
QY 567 CGGCATGGTCACTTCGCTCGCTGGCGCTCGATGAGGACGAGCATCCGGATCCCGCGCG 626
Db 157008 GCAAGAGTGCATGSCCATTTGGCGCGATCAGGCGGTTGCATCCGTATGCGGTCGCG 157067
QY 627 ATGTCGGGACTGTCGGCATGAGGCGACCCACGCGCTGTCGCTCTTACGGCTGAC 686
Db 157068 GTGAGCGGCGATCGTGGGCAAGCGCGCATGACGCTGTGTCCTTACCGGTCGTTT 157127
QY 687 ATACATGACACACCTTTGACACACATCGGCGCCCATCACAGGGGGGTGAGTCAACGC 746
Db 157128 CCGATTGAAGCACCACATGACCATGTGCGCGCGATGGCCAAACAGTGGCGACTCGCG 157187
QY 747 CCGGTCCTCGAGTGTGCGGGCGCGACTGCGGAGCCCTCAGTGGTGGTGAACCT 806
Db 157188 GGTGATCTCGAGTGTGTCGGTGTGTCGCTGTCGCTCTCGAC-----TCAAGGACAGAA 157241
QY 807 TCCGAGCCGAGAACCTACCGCTCCGCGCTCGCGAGGAGTATCCGCTCTGAGTTCG- 865
Db 157242 TCCGAGCGGTGATGTTGTGTCGGACCTTGATCAAGGCTTTCGCGCTGAAATCGG 157301
QY 866 --CGGTGTCGAGAGTCACTGGAGCCGAACGGTGGACCCCGGACGTGATCCCGCGTT 923
Db 157302 TTTGTCGCGCAAGGTTTCGCCATCCCGCGCATGTCGAGCCGCGAGTGGATCGCTGTT 157361
QY 924 CAACAGGAGTGGCGGGCTCGAGAGCGCGGTGCGACCATCGAGCGGCTCTCGGTGCC 983

Db 157362 CAAGCTGCGGTGGCACAACTGCAAGCCTCGGGGCCACGGTGGCGAGGCAAGCGTGC 157421
QY 984 GTTGTGAGCGCGCGCTGGCTATCCAGAGCGCGGTGATGGCTTTCAACCGCGCGCTAT 1043
Db 157422 GTGCGACCATGTTGTGGCGCTGGACATGTGGAAACGTATCGCCACTGACCGCGCGGCGTA 157481
QY 1044 GGGGAGCTCCGCGCGGTGTGGGCTTACTTCCACAAAGGGCGCGGTGGAGAGTCAAGCACCGCGCT 1103
Db 157482 CCAGATGCTCAAGGCAATGCTACGGCATGACGTCGACGGCTACTACGACCCAGAGAT 157541
QY 1104 CACGA-----CGCGCGCCAGAGTGCACCAACCAAGAGATCTGCGGATCTCTGCTCCG 1157
Db 157542 CATGAGCTACTTCGGTGCCAAAGCTCGCGAACACGCAACACGCTGTCCAGCAGCGTACG 157601
QY 1158 GCTGATCTGCTGATCCGCGAGCACCTGCGCGACGAATACCTCGGATCCCACTACGCGAA 1217
Db 157602 CGTGTGGCGCTGACCGGCAATTACAGCTGAACACCTGCAAGGTCGCTTACTACGCCAA 157661
QY 1218 GGGCGAACAACCTGGGTGGAGTTCGGCAAGCAGATGACGCGCTCTCCAGGACCGGGC 1277
Db 157662 GCGCGCATGCTGTGCGCGAACTCACTCGCCAATACGAGAAAGCGTTCAAGGACTTCGA 157721
QY 1278 TGCATCTGACCCCGACCGCTACCTGCTGCGCAACGAGCTGTGAGCGCTGTGAGCGTTCGGCAAGA 1337
Db 157722 CGTCTGGTATGCGGACCATGCGTTCGTGGCGACC---CGTTGACCGCGCGCTATG 157778
QY 1338 CACCATGTCCATGATCCACGATGACGGCAATGCGATCCTCAACAGTGCCTGCGGTGGA 1397
Db 157779 GCCGATTGAAGAGTACGTGACAGCGGCTGAACATCTCGCAACACCGCACCGTTTGA 157838
QY 1398 CTTACCGGTACCGCGGTGACGGTGCACCGGTGCGGCGA 1442
Db 157839 CTTACCGGCGACCGCGCACCTCGATCCCGCGCGTCTCGCGA 157883

RESULT 11
REAG Rhodococcus rhodochrous amidase gene.
LOCUS 1958 bp DNA linear BCT 04-FEB-1999
DEFINITION Rhodococcus rhodochrous amidase gene.
ACCESSION D16207
VERSION D16207.1 GI:468490
KEYWORDS amidase.
SOURCE Rhodococcus rhodochrous
ORGANISM Rhodococcus rhodochrous
Bacteria: Actinobacteria; Actinobacteridae; Actinomycetales;
Corynebacterineae; Nocardiaceae; Rhodococcus.
REFERENCE 1 (bases 1 to 1958)
AUTHORS Kobayashi, M., Kameda, H., Nagasawa, T., Nishiyama, M., Horinouchi, S.,
Beppu, T., Yamada, H. and Shimizu, S.
TITLE Amidase coupled with low-molecular-mass nitrile hydratase from
Rhodococcus rhodochrous J1. Sequencing and expression of the gene
and purification and characterization of the gene product
Eur. J. Biochem. 217 (1), 327-336 (1993)
JOURNAL 94039054
MEDLINE 7916690
PUBMED 2 (bases 1 to 1958)
REFERENCE Kobayashi, M.
AUTHORS Direct Submission
TITLE Submitted (06-MAY-1993) Michihiko Kobayashi, Kyoto University,
JOURNAL Department of Agricultural Chemistry, Faculty of Agriculture,
Oiwake-tyo, Kitashirakawa, Sakyo-ku, Kyoto, Kyoto 606, Japan
(Tel: 075-753-6114, Fax: 075-753-6128)
COMMENT Submitted (06-May-1993) to DDBJ by:
Michihiko Kobayashi
Department of Agricultural Chemistry
Faculty of Agriculture, Kyoto University
Kitashirakawa, Oiwakecho
Sakyo-ku, Kyoto 606
Japan
Phone: 075-753-6114
Fax: 075-753-6128.
FEATURES Location/Qualifiers

```
source
1. .1958
/organism="Rhodococcus rhodochrous"
/mol_type="genomic DNA"
/strain="J1"
/db_xref="taxon:1829"
217. .1764
/EC_number="3.5.1.4"
/codon_start=1
/transl_table=11
/product="amidase"
/protein_id="BAA03744.1"
/db_xref="GI:563884"
/translation="MSSLTPNSNQMSALNNHFRGLTTPBELBEPALBETALASSET
VERLYERTAPPEQSRWTSADENPLISAMVTTSISDEDEGLAGRTAVKDNVAV
GPMWNSRTPPTFYDADTVRLLDAGATITKRAVCEDLFCFSGASFTSHQPVNR
PWDSRTTGGSSSGSALVAGSDVMAGDQGSIRIPAAFCIVGHKPTHGLPYT
GAFPIERTIDHLGPMWTRVSDAAALTVLACTDGLDPROTHRIEVPVLYALAEAPAG
LRGVVTEGPTPTVSDAADVNAVTAIDVLRASGLTVEBVISIPLHLDAMVAMVIAE
GAQVMDLGNAYGNTDGFYDELLAHFSRQLBHGSLKTVKLGMVSGRYTSEVGG
KRYAMARQLVPEVRAAYDALARYDLVWFTLYPTATKIPITDIPDLADYDIALSMI
INTAFDVTGHPACVSPADLVHGLPTGMHIIIGKHFDATVLRVAQLYEHAAGVNPVPP
AAAGTLT"
ORIGIN
Query Match 12.7%; Score 200.6; DB 1; Length 1958;
Best Local Similarity 49.8%; Pred No. 5.9e-17;
Matches 649; Conservative 0; Mismatches 639; Indels 15; Gaps 5;
QY 216 ACCGACCGCGGACGAGACCCGTTCAACGCAATTATCCGGTCTTCGCCGGTGGAGGCGC 275
DB 405 ACCACGCGGACGAGAACCCGCTGAGCGCCGTGTATCTACACCACTCGATCAGCGAAAC 464
QY 276 CACGAGGCGGACCTGACGACCTGACCGCGCGATCAAGCATCGATCGCATCGCGCGG 335
DB 465 GACGAAAGCCCCCTCGCGGCGGACGCTGCGGTGAAGACAAACATCGCGAGTCGCGCGG 524
QY 336 TATGCCACACGAAACCGGTCCGAGTCTCCCGATGTCCGATCGCCACCGAGGATCGCGT 395
DB 525 CGTCCGATGATGAACCGCTCCGAAACCGTCCGAGGCTTACCCCGCTACGACGCCAC 584
QY 396 GGTGTGAGGCGCTCTCGCGGAGCGCGCACCATCGTCGGCAGACGAACTCGAGGA 455
DB 585 CGTCGTACCGGACCTGCTCGACGCGCGGCAACCATACCGCGGAAAGCGGTGTGCGAGA 644
QY 456 CATGCCGAT--GGGTATCGGTGAAGGACGCGTCTACGCTCTCGGTGAAACCGGACAA 512
DB 645 TCTCTGCTTCCGCGCGCAGCTTACATTCACACCCCGCGGTTCGCAACCCCTCGGA 704
QY 513 CCCCAGCGACGCGGTGATCTTCAGCGGCTCGCGCTGCGGTGCGTCCGCGAT 572
DB 705 CGAAAGCGGATCACCGCGGCTGCTCCAGCGGACGCGGCGCCCTGTGTCGCGACGCGCA 764
QY 573 GGTGACTTCGCGCTGGGCGTGCATGAGCAGGACGATCGGATCCCGCGCGCATGTG 632
DB 765 GTTGATATGGAGTCCGCGGACGAGCGGCTTCGATCGCATCCCGCGCGGTCTG 824
QY 633 CGGATGTGTCGATGAAGGACCGACCGGCTGTGCGCTTACGCGCTGACATACAT 692
DB 825 CGGCATCGTCGACACAAACCCACCGGACTGTGCTCCCTATACGGGAGCATTTCCCAT 884
QY 693 GGACACACCTTGGACCATCGGCGCATACACGAGGCGGTTCGAGCTCAACGCCCGG 752
DB 885 CGAACGAACCATCGACCATCTCGTCCGATGACGCGCAGCTCAGCGACGCGCGG 944
QY 753 CTTGAGGTGTGTGGCGGCGGACTGCGCGACCGCTCAGTGGGTGCGTAACCTTCGGA 812
DB 945 GCTACCGTCTCGCGCGGACCGCGGCGCTCGATCCCC--GACAGACCCACCGGATCGA 1001
QY 813 GCGGAGACTAGGCTCGCGCTCGGAGGAGGATATCCGCTCTGAGATCGCGGTG 872
DB 1002 ACCGTGGACTACCTCGCGCGCTGCGCGAACCCCGATCGGCTCTGCGGTGGTGGT 1061
QY 873 CGAGGAGTCACTGGAGCCGAAAGGTGCGACGCGGACGTGATCGCGCGGCTTCAACAGG 932
```

```
DB 1062 CACCGAAGGCTTCGACACCCCTGTCTCCGACGCTGCCGTCCGAAATGCCGTCCGACCGC 1121
QY 933 ACTGCGCGCGCTCGAGAGCGCGGTGCGACCATCAGCGGCTCTCGGTGCGCTTGTGAC 992
DB 1122 CATCGACGACTGCGCTGCGCGGACTTACGCTCGAAGAGTCTGATCCCTCGCACT 1181
QY 993 GCGCGCTGCGCTTCCAGAGCGGCTGATGGCTTTCAACGCGCGCTGATGGCGACTC 1052
DB 1182 CGATGCAATGGCGCTCGAAGCTGATCGCCACCGAGGAGCGGCTACCAAGATGCTCGA 1241
QY 1053 CCGCGGTGCGCTACTTCCAAAGGGGCGGTGAGCTCAGCAGCGCGGTGACGACGCGC 1112
DB 1242 CGCAATGCTTACGGATGAACACTGATGGCTTCTACGATCCCGAATCATCGCCACTT 1301
QY 1113 GCGCCAGAGTCCGACCCACCCAC--AAGGATCTGGCGATCTCTGTCCCGCTGATCTG 1169
DB 1302 CTCCCGTCAACGACTCGAGCAGCGTCACCACTGTCGAAGACAGTCAAACTCGTCGG 1361
QY 1170 GATCGCGGAGCACTGCGGCGAGATACCTCGGCATCCACTACGCGAGGCGGACGAC 1229
DB 1362 GTCCGGCGCTACATCGGAGGTAGCGGCGGCAAGTACTACGCCATGGCGCGCAACT 1421
QY 1230 GCGGCTGGAGCTCGGCAAGCAGATCGACGCCGCTCTCCAGGACCGGCTGCACTGCTG 1289
DB 1422 CGTCCCGGAGTCCGCGCGCTACGACGCCCTTGGCTCGGTACGAGCTGCTGTGAT 1481
QY 1290 CCGGACCAAGCTACCGTTGCCAAGCTGTTGAGCGGTGCGGCAAGACCATGTCCAT 1349
DB 1482 GCGCACCTCTCTACACGCGCACCAAGATCCCGAC--ACGGACATTCCTCGGTGCG 1538
QY 1350 GATCCACCGATGACGGGCAATGCGATCTCTCAACAGTCCCGCTGGAGCTCACCGGTCA 1409
DB 1539 CTATCTGGACACCGCACTGTGATGATCATCAACCGCACCATTCGAGCTCACCGTCA 1598
QY 1410 CCGGCGCTGACCGGTGCCACCGGTCGCGGCGAGAGGCGCTGCGCTGGCTGCTCAAGT 1469
DB 1599 CCGCGCTGAGTGTCCCGCTGACCTGCTGTC---CACGGGCTTCCACCGGAATGAT 1655
QY 1470 GATAGCGCGCACTTCGAGGAGTTCGACGCTTACCGCACCGC 1512
DB 1656 CATCGCAAGCATTCGACGACGCGACAGTGTGCGCTGCGC 1698
```

RESULT 12

```
AF315580
LOCUS
DEFINITION
AF315580 9330 bp DNA linear BCT 27-JUL-2001
Agrobacterium tumefaciens hypothermal protein, Lrp-regulator-like
protein, 1-aminocyclopropane carboxylate deaminase-like protein
Bkdr transcription regulator-like protein, enantioselective amidase
(LivH-like protein, and amino acid transport BraF-like protein
genes, complete cds.
AF315580
AF315580.1 GI:13507180
Agrobacterium tumefaciens (Rhizobium radiobacter)
Agrobacterium tumefaciens
Bacteria; Proteobacteria; Alphaproteobacteria; Rhizobiales;
Rhizobiaceae; Rhizobium/Agrobacterium group; Agrobacterium.
1 (bases 1 to 9330)
Trott,S., Bauer,R., Knackmuss,H.J. and Stolz,A.
Genetic and biochemical characterization of an enantioselective
amidase from Agrobacterium tumefaciens strain d3
Microbiology 147 (Pt 7), 1815-1824 (2001)
21322708
PUBMED
11429459
2 (bases 1 to 9330)
Trott,S.
Direct Submission
Submitted (23-OCT-2000) Institut fuer Mikrobiologie, Universitaet
Stuttgart, Allmandring 31, Stuttgart 70569, Germany
```


QY 166 GAACCGCGCAGCGCGCGGATACACGACCGGACATCGCGCGGACACCGGC 225
Db 4073 GAACCAAGCTGCGCGTGA-----AGTATCAACGACGCGTGGCTGCCACACG 4126
QY 226 GACGAAGACCGGTTCAACGCAATTCATCGGTTCTGCCGGTGGAGGGCGCCACGAGGGG 285
Db 4127 CAGGAACACCGGTTGAACGGGTGCTGCGGTTGCGAGATCGAAGGCGCCAAAGGG 4186
QY 286 CCACTGAGCGACCTGACCGCGCGATCAAGACTGCGATCGCGCGGTATGCCAC 345
Db 4187 CGCTCAAGGGGATCGCATAGCCATCAAGACGTCGTCGCGTGGGGTCCCTATG 4246
QY 346 ACGAACGGGTCCCGGATGTCGCCACTGTGATCGCACCGGAGTCCCGTGGTGGAG 405
Db 4247 ATGAATGCTCAAGCTGCTGAAGTTACGTTCCAGAAATCGATGCGACGATTTAACC 4306
QY 406 CGGTGCTCGCGGAGGCGCCACCATCGTCGCAAGACGAACCTCGAGGACATGCGATG 465
Db 4307 CGCATGCTCGATGCGCGGTGCCACGATCGTCGCAAGTCCGGCTCGAGGATTTCCCTTC 4366
QY 466 GGTATCGGTGAAGGCA---GGTCTACGCTCTCGGCTGAACCCGGAACAACCCCGCCAC 522
Db 4367 TCTGCGCGCGCATGACGTGCTGACCGCGCGGTTGTTATCGTAGCTCCACCGCG 4426
QY 523 GGCACGGGTGATCTTCAGCGGCTCCGGCGCTGCGGTCGCTGCGCGCATGGTCACTTC 582
Db 4427 AACCCTCGCGCATCTCGAACGCTCGGCTGTGCTGATATCGACGGGACAAAGTCGATCTC 4486
QY 583 GCCTCGGGGTGATGAGCAGCAGCATCGGATCCCGCGCGCATGTCGCGACTGGTC 642
Db 4487 GTATCGGTGGACACAGCGGCTCATCGGCTACCGTCTGCGTGGTGGCGGTCTAT 4546
QY 643 GGCATGAAGGCAACCAACCGCTGTGCGCTTTACGGCTTGACATACATGACCAAC 702
Db 4547 GGTCTGAAGCCACCTACGCTCTGCTGCTCCCTATACAGCTCGCGCATGATTGAAGGCAC 4606
QY 703 TTGACCAATCGCGCCATCACAGGGGTCGAGCTCAACCGCGCGGCTCTCGAGTG 762
Db 4607 CTGACCATGTCGCGCGATGCGAGCTCGCCCAAGGGCATCGCAAACTGCTAGTGT 4666
QY 763 TTGGCGGGCGGCTGCGCGACCC---TCAGTGGGTGCGTAACCTTCCGGAGCGGA- 818
Db 4667 ATCGCAGGTACGATGCTGACGATCCGCGCCAGCAGGCGCGCATCGTCCCGGCTTCGAC 4736
QY 819 --GAACACGCTCGCGCTCGCGAGGAGTATCGGCTGCTGAGATTCGCGTCTCGAG 876
Db 4727 ACGAACTATCTCCCGCGCTGAGCGTGGCGTGAAGGGGATGAAGATCGCATTTCTTAA 4786
QY 877 GAGTCACCTGAGCCGAACGG-----TGCGACCGCGGAGCTGATCGCC 918
Db 4787 GAAGGTTTGGCATGACGCTCGAGCGCATGCTTCTCGATCCACTTGTGATGAC 4846
QY 919 GCGTTCAACAGGACTGCGCGCTCGAGAGCGCGGTGCGACCATCGAGCGGCTCG 978
Db 4847 TCGTGCCTCGCGCATGAGACGTTTCGCGGCTTGGCGCGAAGTCCGAGAGTCTCG 4906
QY 979 GTGCGCTTGGACGCGCGCTCGGCTATCCAGAGCGCGGTGATGGCTTTCAACCGCGCG 1038
Db 4907 ATACCTGAACATCTGACGCGCTGGCATATCTGACAGCATTTGGCTTGAAGGCTTACC 4966
QY 1039 GCTATCGCGACTCCGC-----CGGTGGGCTACTTCCAGAGGCGCGGTGAGCTAGC 1095
Db 4967 GCTTCGCGGTGAACGGTAACGGTGTGCGAACAACCTGGAACGGGTGTTAACAACGTCG 5026
QY 1096 ACGCGCTCACGAGCGCGCCAGAGTTCGACACCCACCAAGATCTGCGCATCTGTGCC 1155
Db 5027 ATGGCTGATGACTTGTCTCGCGCATGAGAGTTCGCGCTATGACATGCCGCGACCGTT 5086
QY 1156 CGGCTGATGCTGTGATCGCGAGCACCTGCGCGAGCAATACCTCGGATCCACTACGCG 1215
Db 5087 CGTCTGTCTGATCGCGCGAGTATTTCCGCGCTACTACCAATAATCGCTACTACGCG 5146

QY 1216 AAGGCGCAACCTGCGGCTGGAGCTCGGACGAGATCGACCGCTCTCCAGGACCGG 1275
Db 5147 AAGGCGCAACCAAGCTCACTCACTCAATGAGCGTATGACAGGTCTTTGTCGAATAC 5206
QY 1276 GTGCACTCTGACCCCGACCGCTACCGTTGTCACAGCTGTGAGCGGTGGGCAA 1335
Db 5207 GATATCATTTGTGCGCGAGCATCTCGCTGCTCCACCAAGATGCTGAC---CGGGAT 5263
QY 1336 GACACCATGTCATGATCCACGAGTACGCGGCAATGATCTCAACACGTCGCGCTG 1395
Db 5264 GCGGTACCTCGATACGCTGATCGGTGATCAACGATTGAACAGCTTCGCAACTCTGCTTGC 5323
QY 1396 GACCTCACCGGTCACCCGCGCTGACGGTGCACCGGTGCGGCGAGAGGCGCTGCC 1455
Db 5324 GACCTCACCGGCTCATCTCGATGCTCGGTGCTTGGGACTCGCTG---AGGGGTGCTCT 5380
QY 1456 GTTGGCTCCAGTATAGCGCGCACTTCGAGGATCGACGCTTACCGCACCGCGCC 1515
Db 5381 GTCGGCATGATGCTGACGGCGAAACACITTCGACGACGCAACCTCTATTCTGACGCGCA 5440
QY 1516 GTGATCGAGCGCGCGGCTATGG 1539
Db 5441 GCCTTCGAAGCGCGCGGACTGG 5464

RESULT 13
E12519
LOCUS
DEFINITION Nucleotide sequence of Rhodococcus rhodochrous genomic DNA region
containing amidase and nitrilehydratase genes.
ACCESSION E12519
VERSION E12519.1 GI:3251352
KEYWORDS JP 1997009973-A/3.
SOURCE Rhodococcus rhodochrous
ORGANISM Rhodococcus rhodochrous
Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales;
Corynebacterineae; Nocardiaceae; Rhodococcus.
REFERENCE 1 (bases 1 to 4775)
AUTHORS Aoyama,S. and Yoshida,N.
TITLE NITRILE HYDRATASE GENE AND AMIDASE GENE DERIVED FROM RHODOCOCOCCUS BACTERIUM
JOURNAL Patent: JP 1997009973-A 3 14-JAN-1997;
CHISSO CORP
COMMENT OS Rhodococcus rhodochrous
PN JP 1997009973-A/3
PD 14-JAN-1997 JP 1995184934
PF 27-JUN-1995 JP 1995184934
PI AOYAMA SHIGEKI, YOSHIDA NAOKI
PC C12N15/09,C07H21/04//C12N9/80,C12N9/88,(C12N15/09,C12R1:01),
PC (C12N9/80),
PC C12R1:01),(C12N9/88,C12R1:01);
CC strandedness: Double;
CC topology: Linear;
FH Key
FH Location/Qualifiers
FT source 1..4775
FT /organism='Rhodococcus rhodochrous' FT
FT /strain='IFO15564',
FT RBS 330..334
FT CDS 345..1910
FT /product='amidase'
FT RBS 1973..1977
FT CDS 1984..2607
FT /product='nitrilehydratase alpha subunit' FT
RBS 2622..2627
FT CDS 2634..3272
FT /product='nitrilehydratase beta subunit' FT
FT terminator 3288..3329
FT CDS 3371..4570
FT Location/Qualifiers
FT /product='unknown protein'.
FT 1..4775
FT /organism='Rhodococcus rhodochrous'

/mol_type="genomic DNA"
/db_xref="taxon:1829"

ORIGIN

Query Match 11.8%; Score 186.4; DB 6; Length 4775;
Best Local Similarity 47.8%; Pred. No. 3.6e-15;
Matches 627; Conservative 0; Mismatches 671; Indels 15; Gaps 5;
QY 217 CCGACCGGCGAGAACCGGTTCAACGATTCATCCCGTTCGCGGGTGGAGGGCC 276
DB 534 CCAACAGCGAGGAAATCTTTGAGCGTTGGTATGTGACCAAGCATCCGCCGACG 593
QY 277 ACGAGGGCCACTAGAGCACTGACCGCGCGATCAAGGACTGATGCCATCCCGGT 336
DB 594 TCGGACGGCGTCTTACCGCGCGACCGCGTGGCGATCAAGGCAACGTGACCGTGGCGCGA 653
QY 337 ATGCCACACCAACCGGGTCCGGATGCTCCCGACTGTGATGCCACCGAGGATGCCGTG 396
DB 654 GTTCCGATGATGAACGGGTCTCGGACAGTAGAGGGTTCACTCGTCTCGGACGGACT 713
QY 397 GTGGTGGAGCGCTGCTCGCGCAGCGGCCACCATCGTCGGCAAGAACCTCGAGGAC 456
DB 714 GTGATCACTCGACTACTGCGCGCGGTGCAACCGTCCGGGCAAGAGCTGTGTGAGGAC 773
QY 457 ATGGCGAT---GGGTATCGTGAAGCAGCGCTACGGTCTGCGCTGAAACCGGAACAC 513
DB 774 CTGTGTTCTCGGTTCGAGCTTCAACCGGCAAGCGGACCGTCCGCAATCCATGGAC 833
QY 514 CCGGCCACCGGACCGGTGGATCTTCCAGCGCTCCGGCGTCCCGTCCGTCGCCGATG 573
DB 834 CCACAGCGTGAAGCAGGTGGATCATCCGGTGGCAGTGGCGTCTCGTCCGAACCGTGAC 893
QY 574 GTGCACTTCGCGCTCGGCTCGATGAGGAGCAGCAGCATCCGATCCCGCGCATGTGTC 633
DB 894 GTGATTTGGCATCGCGCGGATCAGGTGGATCGATCCGATCCCGCGGCAATTCGTC 953
QY 634 GGACTGTCGCGATGAAGCGACCCACCGCGCTGGTCCGCTTTACGGCTGCATACATCATG 693
DB 954 GCGTGTGTCGACAAAGCGAGTTCGGGTCTCGTCCGATACCGGTGCAATTCCTCCATC 1013
QY 694 GACCACACTTGGACACATCGGCGCCATCAGGAGGGGTGAGTCAAGCGCGGTC 753
DB 1014 GAGCGAACCAATCGACCATCTCGCGCCGATCAACGCAACCGTCCAGATGCCCACTGATG 1073
QY 754 CTCGAGTGTGCGGGGCGGACTGGCGGACCTTCAGTGGGTGGTAACCTTCGGAG 813
DB 1074 CTCCTGTCATCGCGGTTCGAGCGTACGACCGTACGACCGCAAGCGGACAGGT---CGAA 1130
QY 814 CCGGAGAACTACGGCTTCGCGCTCGCGGAGGAGTATCCGGTCTGAGATCCGCTGCTC 873
DB 1131 GCAGGTGACTATCTGTCCACCTTCGACTCCGATGTGGATGGTCTGCGAATCGGGATCGTT 1190
QY 874 GAGGAGTCACTGAGCGGAACGTGCGCGGACGTGATCCCGGTTCAACGAGGA 933
DB 1191 CGAAGAGTTTCGGGACCGGCTCTACAGCCGAGGTGACAGCCGATCCGCGAGCG 1250
QY 934 CTGCGCGCGCTCGAGACGCGCGGTGCGACCATACGACGGGTCTCGGTGCGCTTGGAGCG 993
DB 1251 GCACACAGTCTGTCCGAAATCGGTTGCGACCGTGAAGGAAATCAATCCGTTGGCACTTA 1310
QY 994 GCGGCTGGCTATCCAGAGCG---GGGTATGGCTTTCAACGCGCGCTATGGCGGAC 1050
DB 1311 CATGCTTTCCACATCTGGAACGTTGATCGCCACGAGCGGTGTGCTTACAGATGTTGGAC 1370
QY 1051 TCGCGCGGTGGGCTACTTCCAAAGGGGCGGTGGAGCTCAGCACCGCCCGTCAAGCG 1110
DB 1371 GGCAACGGAACGCGATGAACGCGGAAGGTTTGTACGATCCGNNNNNNNNNNNNNNNN 1430
QY 1111 GCGGCCAGAGTCGACACCCACAGGATCTGGCATCTGTCCCGGTGATGCTGTG 1170
DB 1431 NNN 1490
QY 1171 ATCGCGGAGCACCTGCGCGAGCAATACCTCGCATCCATACGCAAGGCGCAGAACCTG 1230

DB 1491 ACCGGCACACCGGCATCACACCTTCGCGCGCGAGGTACGGCAAGCCCGAACCTC 1550
QY 1231 CGGCTGAGCTCGGCAAGCAGATCGACGCGTCTCCAGACCGGCTGCACATGCTGAC 1290
DB 1551 GTACCGTCCCGCGCGCTACGACATGCTTTGAGACAATTCAGCTCTCTGGTATG 1610
QY 1291 CCGACCAAGCTACCTGTTCCAAAGAGCTGTTGAGCGGTGGCAAGACACCATGTCCATG 1350
DB 1611 CCCACATGCGCTACGTGCGATCCGAACCTACCGCGCAAGGACGCTGATGTCGAACCTTC 1670
QY 1351 ATCCCGGATGAGGGCAATGCGATCCTCAACACGTCGCCGCTGACCTCACCGGTGTC 1410
DB 1671 ATCAC---GAAGGCTCTCGGATGTCGCCACACAGACCGTTCGACGTGACCGGACAT 1727
QY 1411 CCGCGCTGACGTCGCCCGGTCGCGGCGAGAGGCGCTGCCCTGTGCGCTCCCAAGTG 1470
DB 1728 CCGTCCGTGCTGCTCCGCGCGCTG---GTGAACGGGCTTCGGTTCGGAATGATGATC 1784
QY 1471 ATAGGCGCCACTTCGAGGATGAGCTCTACCGCACCGGCGCGCTGATCGA 1523
DB 1785 ACCGGCAAGACCTTCGACCATGCGAGCTCCTCCGGTTCGGCGCGCATTCGA 1837

RESULT 14

AE016925/c

LOCUS

DEFINITION

Chromobacterium violaceum ATCC 12472 section 16 of 16 of the

complete genome.

ACCESSION

AE016925 AE016825

VERSION

AE016925.1 GI:34105514

KEYWORDS

SOURCE

ORGANISM

Chromobacterium violaceum ATCC 12472

Chromobacterium violaceum ATCC 12472

Bacteria; Proteobacteria; Betaproteobacteria; Neisseriales;

Neisseriaceae; Chromobacterium.

1 (bases 1 to 208524)

Brazilian National Genome Project Consortium

The complete genome sequence of Chromobacterium violaceum reveals

remarkable and exploitable bacterial adaptability

Proc. Natl. Acad. Sci. U.S.A. 100 (20), 11660-11665 (2003)

22882880

PUBMED

14500782

2 (bases 1 to 208524)

Vasconcelos A.T.R., de Almeida, D.F., Almeida, P.C., de

Almeida, L.G.P., de Almeida, R., Gonçalves, J.A.A., Andrade, E.M.,

Antonio, R.V., Araujo, J., de Araujo, M.F.P., Filho, S.A., Azevedo, V.,

Batista, A.J., Bataus, L.A.M., Batista, J.S., Belo, A., vander Berg, C.,

Blamey, J., Bogo, M., Bonato, S., Bordignon, J., Brico, C.A.,

Brocchi, M., Burity, H.A., Camargo, A.A., Cardoso, D.D.P.,

Carneiro, N.P., Carraro, D.M., Carvalho, C.M.B., Cascardo, J.C.M.,

Cavada, B.S., Chueite, L.M.O., Pasa, T.B.C., Duran, N., Fagundes, N.,

Falcao, C.L., Fantinatti, F., Farias, I.P., Felipe, M.S.S.,

Ferrari, L.P., Ferro, J.A., Ferro, M.I.T., Franco, G.R.,

Freitas, N.S.A., Furlan, L.R., Gazinelli, R.T., Gomes, E.A.,

Gonçalves, P.R., Grangeiro, T.B., Grattapaglia, D., Grissard, E.C.,

Guimarães, C.T., Hanna, E.S., Hungria, M., Jardim, S.N., Laurino, J.,

Leoi, L.C.T., Fassarella, L., Lima, A., Loureiro, M.F., Lyra, M.C.P.,

Macedo, M., Madeira, H.M.F., Manfio, G.P., Maranhão, A.Q.,

Martins, W.S., di Mauro, S.M.Z., de Medeiros, S.R.B., Meissner, R.V.,

Menck, C.F.M., Moreira, M.A.M., Nascimento, F.F., Nicolas, M.F.,

Oliveira, J.G., Oliveira, S.C., Paixão, R.F.C., Parente, J.A.,

Pedrosa, F.O., Pena, S.J.D., Pereira, J.O., Perleira, M.,

Pinto, L.S.R.C., Pinto, L.S., Porto, J.I.R., Potrich, D.P.,

Neto, C.E.R., Reis, A.M.M., Rigo, L.U., Rondinelli, E., dos

Santos, E.B.P., Santos, F.R., Schneider, M.P.C., Seuneh, H.N.,

Silva, A.M.R., da Silva, A.L.C., Silva, D.W., Souza, E.M., Simoes, I.C.,

Souza, R.C., Steffens, M.B.R., Steindel, M., Teixeira, S.R.,

Urmey, T., Vetcore, A., Wassem, R., Zaha, A. and Simpson, A.J.G.

Direct Submission

Submitted (22-JAN-2003) Labinfo, LNCC - Laboratório Nacional de

FEATURES source	Computacao Cientifica, Rua Getulio Vargas 333, Petropolis, RJ 25651070, Brazil	
	Location/Qualifiers	
	1..208524	
	/organism="Chromobacterium violaceum ATCC 12472"	
	/mol_type="genomic DNA"	
	/strain="ATCC 12472"	
	/db_xref="taxon:243365"	
	complement(96..563)	
	/gene="pilB2"	
	/locus_tag="CV4209"	
gene	complement(96..563)	
	/gene="pilB2"	
	/locus_tag="CV4209"	
	/note="identified by sequence similarity; putative; ORF located using GeneMark/Blastx/COG2165"	
	/codon_start=1	
	/transl_table=11	
	/product="type IV pilin"	
	/protein_id="AAQ61869.1"	
	/db_xref="GI:34105515"	
	/translation="MKQKQKGGFTLIELMIVAVIGILAAIIPAYQDYTKRVRSE GLADAAKTATYATYATNNAWAGTAPFTNYGLAASITGSSVINVLAVGVQIA YNTVSGALVDLVPVTSAGSVQVYCTGTGTTAIASQLSANWVSTCRQ"	
CDS	838..2184	
	/gene="bioA"	
	/locus_tag="CV4210"	
	838..2184	
	/gene="bioA"	
	/locus_tag="CV4210"	
	/EC_number="2.6.1.62"	
	/note="identified by sequence similarity; putative; ORF located using GeneMark/Blastx/COG0161"	
	/codon_start=1	
	/transl_table=11	
gene	/product="adenosylmethionine-8-amino-7-oxononanoate transaminase"	
	/protein_id="AAQ61870.1"	
	/db_xref="GI:34105516"	
	/translation="MSNQAWLERSFAAVWHPTQMKHERLPIVPIASADGVMLTDPD GKRYLDGVSSWVNLFGHGHPRIKOAVRKOLDLSLHVMLAGFTHRPVLSERLAALS GLGHAFYSGDASATEIALKMSFHYKXNVGEPKOTRFVSLNSVHGTVGALAVTDVP LFSSTYALLPFGHRAQSPDARQAVGERAADVALRAARSENLLAKHGGEIAAIVE PLVQAAGAHGHDPIYTELRRLCDQYOVHLIAEIAVGFRTGSGFFAYQQAQITPDF LCLSGKITGFLPSVLTDRDEIYQAFYHDDVARGFLSHSYTGNALCAALAVLDI FDEVDLVAEVAEGFTQWRPIAERPERHFRQGMWAFDVTARSDFALAYFSA MLNRGLLRPTGKSVYFPPYTLKSEEMDLASVDADEVLAGESVGGDPTALP"	
	2347..3816	
	/locus_tag="CV4211"	
	2347..3816	
	/locus_tag="CV4211"	
	/note="identified by sequence similarity; putative; ORF located using Glimmer/Blastx"	
	/codon_start=1	
CDS	/transl_table=11	
	/product="conserved hypothetical protein"	
	/protein_id="AAQ61871.1"	
	/db_xref="GI:34105517"	
	/translation="MRQAREPGLVPAGPTDGSRAALQAAQALSEGAPLLSAGERS FARRGRILAAAGHQHQAQVQVPGLOANPARKILPPFAAGATAQAGAAVHDD GGAATPEHQAPKALLSASRLLAGLPVAVAGAAALAGQVAGRRRHAGRIHPAASAR SDRHQVAPRRPALIGMTGLRPGATGDAVAGGPDGPPGLCAGPGRCGCPAALSAGOA LAVPARLPARTGRAGDETPQCTTADQDGRSLSGARESPSPAESGSGMAPAPT CAGPLGRDGGDHRRRHLAGPQRQLAARRRRPREQRPAPARLLLPADCRQPE RIWLPAPAPARLARGAVAGPARPAGRVSVRSALGIDALRQSGSGRRRDAGP AAAGEGHAQHQHPQRQPTGLAARQREPAGAAVAPPSLPQVARVQASGRKRATG PRFQAASAIPFPADGIPPTQRAILTWKRRFAFVTLFSL"	
	3845..4963	
	/gene="proB"	
	/locus_tag="CV4212"	
	3845..4963	
	/gene="proB"	
CDS	/locus_tag="CV4212"	
	/note="identified by sequence similarity; putative; ORF located using Glimmer/Blastx/COG3575"	
	/codon_start=1	
	/transl_table=11	
	/product="holliday junction DNA helicase ruvB"	
	/protein_id="AAQ61875.1"	
	/db_xref="GI:34105521"	
	/translation="MIETDKLFGAAPERRIVTPQASQEBEALRALPKLIDEVYVGQ KXARQELEIFBAKRGESALDLDLVYDAENLSPDREALQORRLASVHPDLPHWVTQAA VHLWFADRFGYPVEPLPSLEAALATWPEYATAVGVALNADDSRLRIAPFGLDLDLFAMT VRRNPASASLEIYRQVAKRYQARMPRTVVLE"	
	complement(7321..8358)	
	/genes="ruvB"	
gene	/locus_tag="CV4215"	
	complement(7321..8358)	
	/genes="ruvB"	
	/locus_tag="CV4215"	
	/note="identified by sequence similarity; putative; ORF located using Blastx"	
	/codon_start=1	
	/transl_table=11	
	/product="holliday junction DNA helicase ruvB"	
	/protein_id="AAQ61875.1"	
	/db_xref="GI:34105521"	
CDS	/translation="MIETDKLFGAAPERRIVTPQASQEBEALRALPKLIDEVYVGQ KXARQELEIFBAKRGESALDLDLVYDAENLSPDREALQORRLASVHPDLPHWVTQAA VHLWFADRFGYPVEPLPSLEAALATWPEYATAVGVALNADDSRLRIAPFGLDLDLFAMT VRRNPASASLEIYRQVAKRYQARMPRTVVLE"	
	complement(7321..8358)	
	/genes="ruvB"	
	/locus_tag="CV4215"	
	complement(7321..8358)	
	/genes="ruvB"	
	/locus_tag="CV4215"	
	/note="identified by sequence similarity; putative; ORF located using Glimmer/Blastx/COG3575"	
	/codon_start=1	
	/transl_table=11	
CDS	/product="polyamine ABC transporter system, substrate-binding protein"	
	/protein_id="AAQ61873.1"	
	/db_xref="GI:34105519"	
	/translation="NNKIALAALAVACGSAAEVDLHIYNNVALSPDTAKRFQSC KCKLVQDYVDNEEMAKIALAGAKGYDVFPTAFVNTLLKQGLQPLDKSLPLNKN LNSGVIALNQPDPCGNRYAATPVVSLTILGYNATOLKAGVKGKSNALVDFDALLA KIKGVTVLDSQBELMAALVGLGDANSTPADKAAETIRKAKPYKAAFNQSYI KELTVNHWALGYNDLFOAQODAKNAKRFELAYRQOEGNVLAIDNMILLKDAPR POLAHKFINFLDGRNASEISIGATNPVKAEEAFFKPIKANPVMIDPSKGIYA LKLDVYGRRELNLWSQVKTGR"	
	complement(6173..6757)	
	/locus_tag="CV4214"	
	complement(6173..6757)	
	/locus_tag="CV4214"	
	/note="identified by sequence similarity; putative; ORF located using Glimmer/Blastx/COG3575"	
	/codon_start=1	
gene	/transl_table=11	
	/product="conserved hypothetical protein"	
	/protein_id="AAQ61874.1"	
	/db_xref="GI:34105520"	
	/translation="WPSGRLEQQQLSGLVSLWALRAARALGLSDWCIGAGAVR GLVWDALHGFASPEPSALSDDLVYDAENLSPDREALQORRLASVHPDLPHWVTQAA VHLWFADRFGYPVEPLPSLEAALATWPEYATAVGVALNADDSRLRIAPFGLDLDLFAMT VRRNPASASLEIYRQVAKRYQARMPRTVVLE"	
	complement(7321..8358)	
	/genes="ruvB"	
	/locus_tag="CV4215"	
	complement(7321..8358)	
	/genes="ruvB"	
CDS	/locus_tag="CV4215"	
	complement(7321..8358)	
	/genes="ruvB"	
	/locus_tag="CV4215"	
	/note="identified by sequence similarity; putative; ORF located using Blastx"	
	/codon_start=1	
	/transl_table=11	
	/product="holliday junction DNA helicase ruvB"	
	/protein_id="AAQ61875.1"	
	/db_xref="GI:34105521"	
gene	/translation="MIETDKLFGAAPERRIVTPQASQEBEALRALPKLIDEVYVGQ KXARQELEIFBAKRGESALDLDLVYDAENLSPDREALQORRLASVHPDLPHWVTQAA VHLWFADRFGYPVEPLPSLEAALATWPEYATAVGVALNADDSRLRIAPFGLDLDLFAMT VRRNPASASLEIYRQVAKRYQARMPRTVVLE"	
	complement(7321..8358)	
	/genes="ruvB"	
	/locus_tag="CV4215"	
	complement(7321..8358)	
	/genes="ruvB"	
	/locus_tag="CV4215"	
	/note="identified by sequence similarity; putative; ORF located using Glimmer/Blastx/COG3575"	
	/codon_start=1	
	/transl_table=11	
CDS	/product="polyamine ABC transporter system, substrate-binding protein"	
	/protein_id="AAQ61873.1"	
	/db_xref="GI:34105519"	
	/translation="NNKIALAALAVACGSAAEVDLHIYNNVALSPDTAKRFQSC KCKLVQDYVDNEEMAKIALAGAKGYDVFPTAFVNTLLKQGLQPLDKSLPLNKN LNSGVIALNQPDPCGNRYAATPVVSLTILGYNATOLKAGVKGKSNALVDFDALLA KIKGVTVLDSQBELMAALVGLGDANSTPADKAAETIRKAKPYKAAFNQSYI KELTVNHWALGYNDLFOAQODAKNAKRFELAYRQOEGNVLAIDNMILLKDAPR POLAHKFINFLDGRNASEISIGATNPVKAEEAFFKPIKANPVMIDPSKGIYA LKLDVYGRRELNLWSQVKTGR"	
	complement(6173..6757)	
	/locus_tag="CV4214"	
	complement(6173..6757)	
	/locus_tag="CV4214"	
	/note="identified by sequence similarity; putative; ORF located using Glimmer/Blastx/COG3575"	
	/codon_start=1	
gene	/transl_table=11	
	/product="conserved hypothetical protein"	
	/protein_id="AAQ61874.1"	
	/db_xref="GI:34105520"	
	/translation="WPSGRLEQQQLSGLVSLWALRAARALGLSDWCIGAGAVR GLVWDALHGFASPEPSALSDDLVYDAENLSPDREALQORRLASVHPDLPHWVTQAA VHLWFADRFGYPVEPLPSLEAALATWPEYATAVGVALNADDSRLRIAPFGLDLDLFAMT VRRNPASASLEIYRQVAKRYQARMPRTVVLE"	
	complement(7321..8358)	
	/genes="ruvB"	
	/locus_tag="CV4215"	
	complement(7321..8358)	
	/genes="ruvB"	
CDS	/locus_tag="CV4215"	
	complement(7321..8358)	
	/genes="ruvB"	
	/locus_tag="CV4215"	
	/note="identified by sequence similarity; putative; ORF located using Blastx"	
	/codon_start=1	
	/transl_table=11	
	/product="holliday junction DNA helicase ruvB"	
	/protein_id="AAQ61875.1"	
	/db_xref="GI:34105521"	
gene	/translation="MIETDKLFGAAPERRIVTPQASQEBEALRALPKLIDEVYVGQ KXARQELEIFBAKRGESALDLDLVYDAENLSPDREALQORRLASVHPDLPHWVTQAA VHLWFADRFGYPVEPLPSLEAALATWPEYATAVGVALNADDSRLRIAPFGLDLDLFAMT VRRNPASASLEIYRQVAKRYQARMPRTVVLE"	
	complement(7321..8358)	
	/genes="ruvB"	
	/locus_tag="CV4215"	
	complement(7321..8358)	
	/genes="ruvB"	
	/locus_tag="CV4215"	
	/note="identified by sequence similarity; putative; ORF located using Glimmer/Blastx/COG3575"	
	/codon_start=1	
	/transl_table=11	
CDS	/product="polyamine ABC transporter system, substrate-binding protein"	
	/protein_id="AAQ61873.1"	
	/db_xref="GI:34105519"	
	/translation="NNKIALAALAVACGSAAEVDLHIYNNVALSPDTAKRFQSC KCKLVQDYVDNEEMAKIALAGAKGYDVFPTAFVNTLLKQGLQPLDKSLPLNKN LNSGVIALNQPDPCGNRYAATPVVSLTILGYNATOLKAGVKGKSNALVDFDALLA KIKGVTVLDSQBELMAALVGLGDANSTPADKAAETIRKAKPYKAAFNQSYI KELTVNHWALGYNDLFOAQODAKNAKRFELAYRQOEGNVLAIDNMILLKDAPR POLAHKFINFLDGRNASEISIGATNPVKAEEAFFKPIKANPVMIDPSKGIYA LKLDVYGRRELNLWSQVKTGR"	
	complement(6173..6757)	
	/locus_tag="CV4214"	
	complement(6173..6757)	
	/locus_tag="CV4214"	
	/note="identified by sequence similarity; putative; ORF located using Glimmer/Blastx/COG3575"	
	/codon_start=1	
gene	/transl_table=11	
	/product="conserved hypothetical protein"	
	/protein_id="AAQ61874.1"	
	/db_xref="GI:34105520"	
	/translation="WPSGRLEQQQLSGLVSLWALRAARALGLSDWCIGAGAVR GLVWDALHGFASPEPSALSDDLVYDAENLSPDREALQORRLASVHPDLPHWVTQAA VHLWFADRFGYPVEPLPSLEAALATWPEYATAVGVALNADDSRLRIAPFGLDLDLFAMT VRRNPASASLEIYRQVAKRYQARMPRTVVLE"	
	complement(7321..8358)	
	/genes="ruvB"	
	/locus_tag="CV4215"	
	complement(7321..8358)	
	/genes="ruvB"	
CDS	/locus_tag="CV4215"	
	complement(7321..8358)	
	/genes="ruvB"	
	/locus_tag="CV4215"	
	/note="identified by sequence similarity; putative; ORF located using Blastx"	
	/codon_start=1	
	/transl_table=11	
	/product="holliday junction DNA helicase ruvB"	
	/protein_id="AAQ61875.1"	
	/db_xref="GI:34105521"	
gene	/translation="MIETDKLFGAAPERRIVTPQASQEBEALRALPKLIDEVYVGQ KXARQELEIFBAKRGESALDLDLVYDAENLSPDREALQORRLASVHPDLPHWVTQAA VHLWFADRFGYPVEPLPSLEAALATWPEYATAVGVALNADDSRLRIAPFGLDLDLFAMT VRRNPASASLEIYRQVAKRYQARMPRTVVLE"	
	complement(7321..8358)	
	/genes="ruvB"	
	/locus_tag="CV4215"	
	complement(7321..8358)	
	/genes="ruvB"	
	/locus_tag="CV4215"	
	/note="identified by sequence similarity; putative; ORF located using Glimmer/Blastx/COG3575"	
	/codon_start=1	
	/transl_table=11	
CDS	/product="polyamine ABC transporter system, substrate-binding protein"	
	/protein_id="AAQ61873.1"	
	/db_xref="GI:34105519"	
	/translation="NNKIALAALAVACGSAAEVDLHIYNNVALSPDTAKRFQSC KCKLVQDYVDNEEMAKIALAGAKGYDVFPTAFVNTLLKQGLQPLDKSLPLNKN LNSGVIALNQPDPCGNRYAATPVVSLTILGYNATOLKAGVKGKSNALVDFDALLA KIKGVTVLDSQBELMAALVGLGDANSTPADKAAETIRKAKPYKAAFNQSYI KELTVNHWALGYNDLFOAQODAKNAKRFELAYRQOEGNVLAIDNMILLKDAPR POLAHKFINFLDGRNASEISIGATNPVKAEEAFFKPIKANPVMIDPSKGIYA LKLDVYGRRELNLWSQVKTGR"	
	complement(6173..6757)	
	/locus_tag="CV4214"	
	complement(6173..6757)	
	/locus_tag="CV4214"	
	/note="identified by sequence similarity; putative; ORF located using Glimmer/Blastx/COG3575"	
	/codon_start=1	
gene	/transl_table=11	
	/product="conserved hypothetical protein"	
	/protein_id="AAQ61874.1"	
	/db_xref="GI:34105520"	
	/translation="WPSGRLEQQQLSGLVSLWALRAARALGLSDWCIGAGAVR GLVWDALHGFASPEPSALSDDLVYDAENLSPDREALQORRLASVHPDLPHWVTQAA VHLWFADRFGYPVEPLPSLEAALATWPEYATAVGVALNADDSRLRIAPFGLDLDLFAMT VRRNPASASLEIYRQVAKRYQARMPRTVVLE"	
	complement(7321..8358)	
	/genes="ruvB"	
	/locus_tag="CV4215"	
	complement(7321..8358)	
	/genes="ruvB"	
CDS	/locus_tag="CV4215"	
	complement(7321..8358)	
	/genes="ruvB"	
	/locus_tag="CV4215"	
	/note="identified by sequence similarity; putative; ORF located using Blastx"	
	/codon_start=1	
	/transl_table=11	
	/product="holliday junction DNA helicase ruvB"	
	/protein_id="AAQ61875.1"	
	/db_xref="GI:34105521"	
gene	/translation="MIETDKLFGAAPERRIVTPQASQEBEALRALPKLIDEVYVGQ KXARQELEIFBAKRGESALDLDLVYDAENLSPDREALQORRLASVHPDLPHWVTQAA VHLWFADRFGYPVEPLPSLEAALATWPEYATAVGVALNADDSRLRIAPFGLDLDLFAMT VRRNPASASLEIYRQVAKRYQARMPRTVVLE"	
	complement(7321..8358)	
	/genes="ruvB"	
	/locus_tag="CV4215"	
	complement(7321..8358)	
	/genes="ruvB"	
	/locus_tag="CV4215"	
	/note="identified by sequence similarity; putative; ORF located using Glimmer/Blastx/COG3575"	
	/codon_start=1	
	/transl_table=11	
CDS	/product="polyamine ABC transporter system, substrate-binding protein"	
	/protein_id="AAQ61873.1"	
	/db_xref="GI:34105519"	
	/translation="NNKIALAALAVACGSAAEVDLHIYNNVALSPDTAKRFQSC KCKLVQDYVDNEEMAKIALAGAKGYDVFPTAFVNTLLKQGLQPLDKSLPLNKN LNSGVIALNQPDPCGNRYAATPVVSLTILGYNATOLKAGVKGKSNALVDFDALLA KIKGVTVLDSQBELMAALVGLGDANSTPADKAAETIRKAKPYKAAFNQSYI KELTVNHWALGYNDLFOAQODAKNAKRFELAYRQOEGNVLAIDNMILLKDAPR POLAHKFINFLDGRNASEISIGATNPVKAEEAFFKPIKANPVMIDPSKGIYA LKLDVYGRRELNLWSQVKTGR"	
	complement(6173..6757)	
	/locus_tag="CV4214"	
	complement(6173..6757)	
	/locus_tag="CV4214"	
	/note="identified by sequence similarity; putative; ORF located using Glimmer/Blastx/COG3575"	
	/codon_start=1	
gene	/transl_table=11	
	/product="conserved hypothetical protein"	
	/protein_id="AAQ61874.1"	
	/db_xref="GI:34105520"	
	/translation="WPSGRLEQQQLSGLVSLWALRAARALGLSDWCIGAGAVR GLVWDALHGFASPEPSALSDDLVYDAENLSPDREALQORRLASVHPDLPHWVTQAA VHLWFADRFGYPVEPLPSLEAALATWPEYATAVGVALNADDSRLRIAPFGLDLDLFAMT VRRNPASASLEIYRQVAKRYQARMPRTVVLE"	
	complement(7321..8358)	
	/genes="ruvB"	
	/locus_tag="CV4215"	
	complement(7321..8358)	
	/genes="ruvB"	
CDS	/locus_tag="CV4215"	
	complement(7321..8358)	
	/genes="ruvB"	
	/locus_tag="CV4215"	
	/note="identified by sequence similarity; putative; ORF located using Blastx"	
	/codon_start=1	
	/transl_table=11	
	/product="holliday junction DNA helicase ruvB"	
	/protein_id="AAQ61875.1"	
	/db_xref="GI:34105521"	
gene	/translation="MIETDKLFGAAPERRIVTPQASQEBEALRALPKLIDEVYVGQ KXARQELEIFBAKRGESALDLDLVYDAENLSPDREALQORRLASVHPDLPHWVTQAA VHLWFADRFGYPVEPLPSLEAALATWPEYATAVGVALNADDSRLRIAPFGLDLDLFAMT VRRNPASASLEIYRQVAKRYQARMPRTVVLE"	
	complement(7321..8358)	
	/genes="ruvB"	
	/locus_tag="CV4215"	
	complement(7321..8358)	
	/genes="ruvB"	
	/locus_tag="CV4215"	
	/note="identified by sequence similarity; putative; ORF located using Glimmer/Blastx/COG3575"	
	/codon_start=1	
	/transl_table=11	
CDS	/product="polyamine ABC transporter system, substrate-binding protein"	
	/protein_id="AAQ61873.1"	
	/db_xref="GI:34105519"	
	/translation="NNKIALAALAVACGSAAEVDLHIYNNVALSPDTAKRFQSC KCKLVQDYVDNEEMAKIALAGAKGYDVFPTAFVNTLLKQGLQPLDKSLPLNKN LNSGVIALNQPDPCGNRYAATPVVSLTILGYNATOLKAGVKGKSNALVDFDALLA KIKGVTVLDSQBELMAALVGLGDANSTPADKAAETIRKAKPYKAAFNQSYI KELTVNHWALGYNDLFOAQODAKNAKRFELAYRQOEGNVLAIDNMILLKDAPR POLAHKFINFLDGRNASEISIGATNPVKAEEAFFKPIKANPVMIDPSKGIYA LKLDVYGRRELNLWSQVKTGR"	
	complement(6173..6757)	
	/locus_tag="CV4214"	
	complement(6173..6757)	
	/locus_tag="CV4214"	
	/note="identified by sequence similarity; putative; ORF located using Glimmer/Blastx/COG3575"	
	/codon_start=1	
gene	/transl_table=11	
	/product="conserved hypothetical protein"	
	/protein_id="AAQ61874.1"	
	/db_xref="GI:34105520"	
	/translation="WPSGRLEQQQLSGLVSLWALRAARALGLSDWCIGAGAVR GLVWDALHGFASPEPSALSDDLVYDAENLSPDREALQORRLASVHPDLPHWVTQAA VHLWFADRFGYPVEPLPSLEAALATWPEYATAVGVALNADDSRLRIAPFGLDLDLFAMT VRRNPASASLEIYRQVAKRYQARMPRTVVLE"	
	complement(7321..8358)	
	/genes="ruvB"	
	/locus_tag="CV4215"	
	complement(7321..8358)	
	/genes="ruvB"	
CDS	/locus_tag="CV4215"	
	complement(7321..8358)	
	/genes="ruvB"	
	/locus_tag="CV4215"	
	/note="identified by sequence similarity; putative; ORF located using Blastx"	
	/codon_start=1	
	/transl_table=11	
	/product="holliday junction DNA helicase ruvB"	
	/protein_id="AAQ61875.1"	
	/db_xref="GI:34105521"	
gene	/translation="MIETDKLFGAAPERRIVTPQASQEBEALRALPKLIDEVYVGQ KXARQELEIFBAKRGESALDLDLVYDAENLSPDREALQORRLASVHPDLPHWVTQAA VHLWFADRFGYPVEPLPSLEAALATWPEYATAVGVALNADDSRLRIAPFGLDLDLFAMT VRRNPASASLEIYRQVAKRYQARMPRTVVLE"	
	complement(7321..8358)	
	/genes="ruvB"	
	/locus_tag="CV4215"	
	complement(7321..8358)	
	/genes="ruvB"	
	/locus_tag="CV4215"	
	/note="identified by sequence similarity; putative; ORF located using Glimmer/Blastx/COG3575"	
	/codon_start=1	
	/transl_table=11	
CDS	/product="polyamine ABC transporter system, substrate-binding protein"	
	/protein_id="AAQ61873.1"	
	/db_xref="GI:34105519"	
	/translation="NNKIALAALAVACGSAAEVDLHIYNNVALSPDTAKRFQSC KCKLVQDYVDNEEMAKIALAGAKGYDVFPTAFVNTLLKQGLQPLDKSLPLNKN LNSGVIALNQPDPCGNRYAATPVVSLTILGYNATOLKAGVKGKSNALVDFDALLA KIKGVTVLDSQBELMAALVGLGDANSTPADKAAETIRKAKPYKAAFNQSYI KELTVNHWALGYNDLFOAQODAKNAKRFELAYRQOEGNVLAIDNMILLKDAPR POLAHKFINFLDGRNASEISIGATNPVKAEEAFFKPIKANPVMIDPSKGIYA LKLDVYGRRELNLWSQVKTGR"	
	complement(6173..6757)	
	/locus_tag="CV4214"	
	complement(6173..6757)	
	/locus_tag="CV4214"	
	/note="identified by sequence similarity; putative; ORF located using Glimmer/Blastx/COG3575"	
	/codon_start=1	
gene	/transl_table=11	
	/product="conserved hypothetical protein"	
	/protein_id="AAQ61874.1"	
	/db_xref="GI:34105520"	
	/translation="WPSGRLEQQQLSGLVSLWALRAARALGLSDWCIGAGAVR GLVWDALHGFASPEPSALSDDLVYDAENLSPDREALQORRLASVHPDLPHWVTQAA VHLWFADRFGYPVEPLPSLEAALATWPEYATAVGVALNADDSRLRIAPFGLDLDLFAMT VRRNPASASLEIYRQVAKRYQARMPRTVVLE"	
	complement(7321..8358)	
	/genes="ruvB"	
	/locus_tag="CV4215"	
	complement(7321..8358)	
	/genes="ruvB"	
CDS	/locus_tag="CV4215"	
	complement(7321..8358)	
	/genes="ruvB"	
	/locus_tag="CV4215"	
	/note="identified by sequence similarity; putative; ORF located using Blastx"	
	/codon_start=1	
	/transl_table=11	
	/product="holliday junction DNA helicase ruvB"	
	/protein_id="AAQ61875.1"	
	/db_xref="GI:34105521"	
gene	/translation="MIETDKLFGAAPERRIVTPQASQEBEALRALPKLIDEVYVGQ KXARQELEIFBAKRGESALDLDLVYDAENLSPDREALQORRLASVHPDLPHWVTQAA VHLWFADRFGYPVEPLPSLEAALATWPEYATAVGVALNADDSRLRIAPFGLDLDLFAMT VRRNPASASLEIYRQVAKRYQARMPRTVVLE"	
	complement(7321..8358)	
	/genes="ruvB"	
	/locus_tag="CV4215"	
	complement(7321..8358)	
	/genes="ruvB"	
	/locus_tag="CV4215"	
	/note="identified by sequence similarity; putative; ORF located using Glimmer/Blastx/COG3575"	
	/codon_start=1	
	/transl_table=11	
CDS	/product="polyamine ABC transporter system, substrate-binding protein"	
	/protein_id="AAQ61873.1"	

```
/gene="ribB"
/locus_tag="CV4216"
8726..9358
/gene="ribB"
/locus_tag="CV4216"
/notes="identified by sequence similarity; putative; ORF
located using GeneMark/BLASTx/COG0108"

Query Match      11.8%; Score 186; DB 1; Length 208524;
Best Local Similarity 49.2%; Pred. No. 1.4e-15; Indels 30; Gaps 5;
Matches 650; Conservative 0; Mismatches 640;

QY 183 GAGCGCGTACAGGACCGGACATCGCGCGGAAACCGACCGGCGAGAGACCGGTTCAA 242
DB 154426 GCGCTGAACCGCAGCTCAAGCCATCGTACCGTAGACCGGAAAGACGCTGCGGA 154367
QY 243 CGCATTTATCCGGTTCTGCGGGTGAAGGCGCCAGGAGGCGCACTAGAGCACTGAC 302
DB 154366 GCGCGCGGCGCGAGCGCGCGCTCGCGCGGCGATGCGCGGCTGACCGGGTGGC 154307
QY 303 CGCGCGGATCAAGGATGTCATCGCCATCGCGGTATGCCACACGAAACGGGTCCGGAT 362
DB 154306 GCTGGTGCACAGGACTGTTCTGCCAGCAAGTTTGGAAACCAAGCTCGGCTCCAGAT 154247
QY 363 GCTCCGACTGTGATCGCCACAGAGATGCCGTGTTGGTGGAGCGGCTGCTCGCGGAGG 422
DB 154246 GCTGGAACAATTCGTTTGGCGGTACAGCGCCACGTTGGTGGAGCATGCGCGCGCGCG 154187
QY 423 CGCCACCATCGTCGGAAGACGAACTCGAGGACATGGCGATCGGTTATCGG--TGAAGG 479
DB 154186 CATGTTGACGCTGGCGCGCCAAATATGACGAGTTTCGGATGGGCTCGTCAATGAA 154127
QY 480 CAGCGTCTACGGTCTGCGCTGAACCGGAAACACCGCGCCAGCGCAACGGGTGGATCTTC 539
DB 154126 CTCGTTCTACGGCGCGTGAGAAACCGGCTGGGACCTGAACCGGATTCGCGGGGTTCGTC 154067
QY 540 CAGCGCTCGCGGCTCGCGTCTGCTCGCGCATGCTCGCTTCCGCTCGCGGTGATGA 599
DB 154066 CGCGGCTCGCGCGCGGCTGCGCGCGCTGCGCGCGGTCGCGCGGTCGCGCACCGAC 154007
QY 600 GGAGGAGCATCGGATCCGCGCCCGCATGCTGTCGAGTCTGTCGCGGATGAAGCGCAACCA 659
DB 154006 CGCGGTTTCATCCGCGCGCGGACGACCTTGGCGGTTCACCGGATCAAGCGGACCTA 153947
QY 660 CGCGCTGTCGCTCTACGCGCTGACATACATGACCAACACCTTGGACCATCGGGC 719
DB 153946 CGCGGTGTGTGCGTTTACGGCATGTGTGCGCTACGCTTCTCGCTGACCGGCGCGCC 153887
QY 720 CATCACAGGGGGGTGAGCTCAACCGCGGGTCTCGAGTGTGTTGGCGGGCGGCGACTG 779
DB 153886 GATCGCCACAGCCCGGAGGACTGCGCGCTGATGCTGAACGTGATGCGCGGCTTCGACGC 153827
QY 780 GCGGACCTTCAGTGGTGGTGAACCTTCGAGAGCGGAGAACTAGCGTCCGCGTCCG 839
DB 153826 GCGGACTCCACGAGCTG-----GAGCGCGCAAGGAGAGACTAGCGCGGACCTGAA 153773
QY 840 CGAGGGAGTATCGGTCGTAGATTGCGGCTGCTGAGGAGTCACTGAGCGGAAACGGTGC 899
DB 153772 CCAGTCTTGTCCGACTGAGGTGGCTGCGCCAGGATATCTTGGCGCGGTCTGGA 153713
QY 900 GACCGCGGATCGGATCCGCGGTTCAACAGGAGTCTGGCGGCTCGAGAGCGCGGCTGC 959
DB 153712 GCGCGAGCTGCGCGCGCGCTGCACAAACGCGGTGGCGGAGCTGAAAGAGTGGCGCGCGA 153653
QY 960 GACCATCGAGCGGTCTCGGTGCGGTTGTGAGCGGCGGCTGCGCTATCCAGAGCGCGT 1019
DB 153652 GCGGTGAGATAGCTTGCCTGCAATACGAGCTGTGATCCCGGCTTATTCGTGATCGC 153593
QY 1020 GATGGCTTTCAACCGCGCGCTATGGCGGATCGCGGCTGTTGGGTACTTCCACAGGG 1079
DB 153592 CCGCGCGAGGCTTCGACCAACCTGTCCCGTACGACGCGGTCCGCTACGCGCAC--CG 153536
QY 1080 GCGGTGAGCTGAGCACCGCGCTCAGACGCGGCGCCAGTTCGCGCACCCACCAAGGA 1139
```

```
DB 153535 CGCCAAAGACTACAGGACTTGTTCGACATGTACGAGAAACCGCGCCGAGGCTTCGG 153476
QY 1140 TCTGGCGATCTGTCTCCGGCTGATGTGTGTGATGTCGCGAGACACTGTCGCGACGATACCT 1199
DB 153475 CGACGAGGTCAAGCGCGCATCTCTGTGTGGTAGTATGTG---CTGAGCCACCGTTACTA 153419
QY 1300 CGGCATCCACTACGCGAAGCGGAGAACCTCGCGCTGAGCTGGCAAGCAGATCGACGC 1259
DB 153418 CGAGCCTTACTCTGAGGCGGCAAGATCCGCGCTGATCGCCAAACGACTTCAAGGC 153359
QY 1260 CGTCTCCAGSACCGGCTGCACTGTGACCCCGACACACCTTACCCTTTCGCAACGAGCT 1319
DB 153358 GCGCTTCGAGCAGTGCAGCTGATCTCTGGGCGCGGTGGCGCCGACCCGCTTCAATAT 153299
QY 1320 GTTGAGCGGTGCGCAAGACACCATGTCATGATCCACGATGACGGGCAATGCGATCCT 1379
DB 153298 C-----GGCGAAGTCCGGGATCCGGTGCAGATGATCTGTCCGACAT 153254
QY 1380 CAACACGTGCGCGCTGACCTCACCGTCAACCGGCGCTGACCGTCCCAACGAGTGGCGG 1439
DB 153253 CTACACGCTGTGCTGAACTTGGCGGCTGCGGCGATGAGCGTGGCGCGGCTTCGC 153194
QY 1440 CGAGAGGGCTGCGCTTGGCTTCCAGTGTAGTGGCGGCACTTCGAGGAGTGCAGCT 1499
DB 153193 CGCAACGCGCGCGCTGCGCTGCGATGATCGGCACTTTCGCGAAGCCAGAT 153134

RESULT 15
BD061400 2822 bp DNA linear PAT 27-AUG-2002
LOCUS Rhodococcus bacterium-derived nitrile hydratase gene and amittase
DEFINITION gene.
ACCESSION BD061400
VERSION BD061400.1 GI:22607005
KEYWORDS JP 2001292772-A/2.
SOURCE Rhodococcus sp.
ORGANISM Rhodococcus sp.
Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales;
Corynebacterineae; Nocardiaceae; Rhodococcus.
REFERENCE 1 (bases 1 to 2822)
AUTHORS Kamaike,H. and Aoki,Y.
TITLE Rhodococcus bacterium-derived nitrile hydratase gene and amittase
JOURNAL Patent: JP 2001292772-A 2 23-OCT-2001;
SHOWA DENKO KK
COMMENT OS Rhodococcus sp.
FN JP 2001292772-A/2
PD 23-OCT-2001
PF 10-APR-2000 JP 2000107855
PI HARUMI KAMAIKE,YUJI AOKI
PC C12N15/09,C12N1/15,C12N1/19,C12N1/21,C12N5/10,C12N9/80,C12N9/
PC 88,C12P13/00,
PC C12P13/02//C12N15/09,C12R1.01), (C12N9/80,C12R1.01), (C12N9/88,
PC C12R1.01),
PC C12N15/00,C12N5/00, (C12N15/00,C12R1.01)
CC amittase
PH Key Location/Qualifiers
FT CDS Location/Qualifiers
1..2822
/organism="Rhodococcus sp."
/mol_type="genomic DNA"
/db_xref="taxon:1831"

ORIGIN
Query Match 11.5%; Score 182.2; DB 6; Length 2822;
Best Local Similarity 53.4%; Pred. No. 1.5e-14; Indels 6; Gaps 2;
Matches 429; Conservative 0; Mismatches 368;
QY 216 ACCGACCGCGGACGAGACCGGCTGAGCGCTGCTGAGGAGTTCGCGGTTCTGCGGTTGAGGCGC 275
DB 1282 ACCGACCGCGGACGAGACCGGCTGAGCGCTGCTGAGGAGTTCGCGGTTCTGCGGTTGAGGCGC 1341
```

Search completed: June 27, 2004, 07:56:33
Job time : 6406.97 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: June 27, 2004, 03:11:33 ; Search time 680.85 Seconds
(without alignments)
9846.004 Million cell updates/sec

Title: US-10-658-691-4

Perfect score: 1578

Sequence: 1 gtgcgaccacatcgccatt.....ggcgaccgtgtcggtaggtag 1578

Scoring table: IDENTITY NUC

Gapop 10.0 , Gapext 1.0

Searched: 3373863 seqs, 2124099041 residues

Total number of hits satisfying chosen parameters: 6747726

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : N Geneseq_29Jan04:*

1: Geneseqn1980s:*

2: Geneseqn1990s:*

3: Geneseqn2000s:*

4: Geneseqn2001as:*

5: Geneseqn2001bs:*

6: Geneseqn2002as:*

7: Geneseqn2003as:*

8: Geneseqn2003bs:*

9: Geneseqn2003cs:*

10: Geneseqn2004s:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match %	Length	DB ID	Description
1	182.2	11.5	2822	4 AAH21845	Aah21845 Rhodococ
2	180.8	11.5	1879	2 AAQ11991	Aaq11991 Enantiose
3	180.8	11.5	1879	2 AAQ25616	Aaq25616 Enantiose
4	175.4	11.1	1566	2 AAT82385	Aat82385 Rhodococ
5	175.2	11.1	1479	7 ACA37683	Ac37683 Prokaryot
6	172.8	11.0	110000	4 AAI99683_33	Continuation (34 o
7	172	10.9	1482	7 ACA38250	Ac38250 Prokaryot
8	170.4	10.8	1482	7 ABZ71125	Abz71125 Mycobacte
9	170.4	10.8	1485	7 ACA40772	Ac40772 Prokaryot
10	170.4	10.8	110000	4 AAI99682_33	Continuation (34 o
11	157.4	10.0	4730	2 AAT82387	Aat82387 R. rhodoc
12	152.2	9.6	1488	7 ACA26208	Ac26208 Prokaryot
13	151	9.6	1455	4 AAS54287	Aas54287 Pseudomon
14	151	9.6	1455	7 ACA42618	Ac42618 Prokaryot
15	146.4	9.3	349980	6 ABO81844	Abq81844 Bifidobac
16	141.4	9.0	1488	7 ACA23486	Ac23486 Prokaryot
17	141.4	9.0	1536	7 ACA27226	Ac27226 Prokaryot
18	140.6	8.9	1449	7 ACA3522	Ac3522 Prokaryot
19	133.2	8.4	2438	2 AAT47765	Aat47765 Comamonas
20	130.6	8.3	1038	7 ADA71204	Ada71204 Rice gene
21	125.4	7.9	1398	9 ADE50885	Ade50885 V. parado
22	124.2	7.9	1485	7 ACA25570	Ac25570 Prokaryot
23	123.8	7.8	1449	7 ACA45943	Ac45943 Prokaryot

24	123.4	7.8	110000	4 AAI99682_26	Continuation (27 o
25	123.4	7.8	110000	4 AAI99683_26	Continuation (27 o
26	121	7.7	1494	7 ACA39866	Ac39866 Prokaryot
27	120	7.6	2456	2 AAQ13362	Aaq13362 Gene enco
28	119.4	7.6	1521	7 ACA52610	Ac52610 Prokaryot
29	114.2	7.2	1443	7 ABZ39903	Abz39903 N. gonorr
30	114.2	7.2	1443	7 ACA41226	Ac41226 Prokaryot
31	114.2	7.2	1458	6 ABK73873	Abk73873 Bacillus
32	113.4	7.2	1386	7 ABZ71160	Abz71160 S. muraya
33	113.4	7.2	1446	7 ACA41891	Ac41891 Prokaryot
34	113.4	7.2	36321	7 ABZ71131	Abz71131 Streptomy
35	111.6	7.1	50925	3 AA81487	Aa81487 N. mening
36	111.6	7.1	110000	3 AA81490_13	Continuation (14 o
37	111.6	7.1	349980	3 AAF21610	Aaf21610 Neisseria
38	111	7.0	861	4 AAH27159	Aah27159 Amidase D
39	110.4	7.0	23673	6 ABZ75344	Abz75344 Human R11
40	108.8	6.9	1844	2 AAO38949	Aao38949 C. acidov
41	107.8	6.8	22934	4 AAS59613	Aas59613 Propionib
42	107.8	6.8	22934	7 ACF64542	Acf64542 Propionib
43	106.8	6.8	1395	7 ACA31528	Ac31528 Prokaryot
44	106.6	6.8	7127	6 ABL58898	Ab158898 Ralstonia
45	102	6.5	1621	9 ADD13197	Add13197 C. glutam

ALIGNMENTS

RESULT 1
AAH21845
ID AAH21845 standard; DNA; 2822 BP.
XX
AC AAH21845;
XX
DT 15-AUG-2001 (first entry)
XX
DE Rhodococcus sp. amidase encoding DNA SEQ ID NO:6.
XX
KW Rhodococcus; nitrilase; nitrile hydratase beta subunit; amidase;
KW nitrile hydratase alpha subunit; carboxylic acid; cyanocarboxylic acid;
KW pharmaceutical; agricultural drug; dye; ds.
XX
OS Rhodococcus sp.
XX
FH Key Location/Qualifiers
FT CDS 1094..2491
FT /*tag= a
FT /product= "amidase"

XX WO200130994-A1.
XX 03-MAY-2001.
XX 25-OCT-2000; 2000WO-JP007464.
XX 26-OCT-1999; 99JP-00303212.
XX 26-JAN-2000; 2000JP-00021797.
XX 10-APR-2000; 2000JP-00107855.
(SHOW) SHOWA DENKO KK.
XX Aoki H, Kamachi H;
XX WPI; 2001-308634/32.
XX P-PSDB; AAB98070.
XX Producing carboxylic acids using Rhodococcus to convert cyano groups of
XX nitrile compounds to carboxyl groups.
XX Claim 29; Page 75-80; 89pp; Japanese.
XX The present invention describes a method for producing carboxylic acids,
XX comprising using a microorganism, whose ability to convert cyano groups
XX to amide groups has been removed or reduced, to convert at least one

cyano group of a nitrile compound to a carboxyl group. Also described are: (1) a microorganism for use in the novel method; (2) a plasmid containing DNA for Rhodococcus nitrilase gene; (3) a host transformed by the plasmid of (2); (4) a nitrilase gene, comprising a 1531 base pair sequence (AAH21843), and encoding a 366 residue amino acid sequence (AAB98067); (5) producing nitrilase, comprising culturing the host of (3) under expression conditions, and recovering the polypeptide; (6) a nitrile hydratase gene, comprising a 2822 base pair sequence (AAH21844), encoding a 229 nitrile hydratase beta subunit (AAB98068) and a nitrile hydratase alpha subunit (AAB98069); (7) producing amides using hosts transformed by a 2822 base pair sequence (AAH21845), encoding for Rhodococcus amidase (AAB98070). The method can be used for producing carboxylic acids and cyanocarboxylic acids which are useful in pharmaceuticals, agricultural drugs, dyes and as starting materials for other chemicals

XX
SQ Sequence 2822 BP; 478 A; 1034 C; 869 G; 438 T; 0 U; 3 Other;

Query Match 11.5%; Score 182.2; DB 4; Length 2822;
Best Local Similarity 53.4%; Pred. No. 1.6e-22;
Matches 429; Conservative 0; Mismatches 368; Indels 6; Gaps 2;

QY 216 ACCGACCGGCGAGAGACCGGTCACAGCATTCATCCGGTCTGCGGGTGGAGGCGC 275
Db 1282 ACCACCGGCGAGAGACCGGTCAGCGCTGGTAGTCACCACTCGATCAGCGAAC 1341

QY 276 CACGGAGGGCCACTAGCGAATCCGCGGATCAAGAGTCAAGTATCGCCATCGCCGG 335
Db 1342 CGACGAAGGCGCCCTCGCGGGGGAACGGTCGCGGTGAAGACAACGTCGAGTCGCGG 1401

QY 336 TATGCCACACCAAGCGGGTCCCGGATGCTCCGACTGTCGCGACCGAGGATGCGGT 395
Db 1402 CGTGCGGATGATGAACGGCTCCGACCGTTCGAGGGCTTCACCCCGGTACGAGCCAC 1461

QY 396 GTGTGTGAGCGCTCTCGCGCAGCGCCACCATCGTTCGGCAAGAACCTCGAGGA 455
Db 1462 CGTCGTACCGCGACTCTCGACCGCGCGCAACCATCACCGGCAAGCGGTGTGCGAAGA 1521

QY 456 CATGGCGAT--GGGTATCGGTGAAGCAGCGCTACGGTCTCGGTGTAACCCGACAA 512
Db 1522 TCTCTGCTCTCGCGCGCAGCTTCACTCCACCCCGAGCGGTCCGCAACCCCTGGGA 1581

QY 513 CCCCAGCGCAGCGGCTGATGATGAGCAGCGCTCGCGCTCGCGCTCGCGGAT 572
Db 1582 CGAAGCGCGATCACCGCGGCTGTCGACGCGAGCGCGCTGTCGCGCAGCGGCA 1641

QY 573 GPTCGACTTCGCGCTCGCGTGCATGAGCAGCGCATCGGATCCGCGCGCGATGTG 632
Db 1642 GTGTGATATGGAGTCCGCGCGGACCGAGGCGGTTCGATCCGATCCCGCGCGGTCTG 1701

QY 633 CGGACTGGTTCGCGATGAAGCGGACCCAGCGCTGTGCTGCTTACGGCTGACATACAT 692
Db 1702 CGGCATCGTCGGACACAAACCCACCGACTGTGTCCCTATACGGGAGCATTTCCCAT 1761

QY 693 GACACACCTTGACCATCGGCGCCATCACAGGGGGTTCAGTCAAGCGCGGT 752
Db 1762 CGAACGAACATCGACCATCGTTCGATGACGCGACGTCAGCGCGCGCGCAT 1821

QY 753 CTTCGAGGTGTTGGCGGGCGCGACTTGGCGCGACCTCAGTGGTGGTAACTTCCGGA 812
Db 1822 GCTCACCGTCTCGCGCGCACCGAGCGCTCGATCCCC--GACAGACCCCGGATCGA 1878

QY 813 CGCGGAGACTAGGTCTCGCGTTCGCGAGGAGTATCCGCTCTGAGATTCGCGGTG 872
Db 1879 ACCGTGGAGTACCTCGCGCGCTGCGCGAACCCGATCGGCTCTGCGGTGGTGGT 1938

QY 873 CGAGGAGTCACTGAGACCGAAACGGTTCGACCGCGGACGTGATCGCGGTTCACACGG 932
Db 1939 CACCGAAGGCTTCGACACCCCTGTCTCCAGCGCTCGGCTCGACATCGGTCGACCGC 1998

QY 933 ACTGGCGGCGCTCGAGAGCGCGGTGCGACCATCGAGCGGTCTCGGTGCGCTTGTGAC 992
Db 1999 CATCGCGTACTCGGTTCGCGCGGACTTACCGCTCGAAGAGGTCTCGATCCCTCGCACCT 2058

QY 993 GCGCGCTGCGCTATCCAGAGCG 1015
Db 2059 CGATCGATGCGCGCTCTGGAACG 2081

RESULT 2
ID AAQ11991 standard; DNA; 1879 BP.
XX AAQ11991;
AC AAQ11991;
XX 27-AUG-2003 (revised)
DT 25-MAR-2003 (revised)
DT 30-AUG-1991 (first entry)
XX
DE Enantioselective amidase gene of Brevibacterium.
XX
XX ESA; S(+)-ketoprofen; racemase; ss.
XX
OS Brevibacterium sp.
XX
FH Key Location/Qualifiers
FT RBS 230..234
FT /*tag= b
FT CDS /standard_name= "Shine-Dalgarno sequence"
FT 245..1810
FT /*tag= a
FT /product= "enantioselective amidase"
XX
PN EP433117-A.
XX
PD 19-JUN-1991.
XX
PF 15-NOV-1990; 90EP-00403232.
XX
PR 11-DEC-1989; 89FR-00016332.
XX
PA (RHON) RHON-POULENC SANTE.
PA (PETR/) PETRE D.
PI Petre D, Cerebaleud E, Mayaux JF, Yeh P;
XX WPI; 1991-179908/25.
XX
PT DNA encoding enantio-selective amidase - used for prodn. of the enzyme
PT for enantio-selective hydrolysis of amide(s) to acids.
XX
PS Claim 2; Fig 8; 36pp; English.
XX
CC Probe Sq762 (AAQ11989) and probe Sq918 (AAQ11990) were used to screen
CC genomic Brevibacterium DNA. Both probes hybridised strongly to a 5.4kb
CC PstI fragment. Cloning and restriction analysis of this fragment showed
CC that the gene coding for the enantioselective amidase is located on a
CC 2.3kb BamHI-PstI fragment (oriented in the direction BamHI towards PstI).
CC The fragment was sequenced. See also AAQ11988 and AAQ11992. (Updated on
CC 25-MAR-2003 to correct PA field.) (Updated on 27-AUG-2003 to correct OS
CC field.)
SQ Sequence 1879 BP; 409 A; 612 C; 531 G; 327 T; 0 U; 0 Other;

Query Match 11.5%; Score 180.8; DB 2; Length 1879;
Best Local Similarity 48.1%; Pred. No. 2.8e-22;
Matches 704; Conservative 0; Mismatches 742; Indels 18; Gaps 6;

QY 66 CGCCCGCCACACTTCGACCTCGACGAGGAACTGGCGCCGCGCTCGTTCGGTCTGGC 125
Db 286 CGCAGGCATTAGGATCACTCTGCACAAACAGCCGCTCGAGTGGCGGCACTGAT 345

QY 126 GGAGATGCTGACTGCTTCGACCTGATCGACGAACTACCGCAACCCGCGCGCGAC 185
Db 346 CGACGGAGCACTGGGTCTCTACGCGTCTGCGACCACTGTCGCGGCGAGCGACCCC 405

CC on 25-MAR-2003 to correct PN field.) (Updated on 25-MAR-2003 to correct
CC PA field.) (Updated on 27-AUG-2003 to correct OS field.)
XX
SQ Sequence 1879 BP; 409 A; 612 C; 531 G; 327 T; 0 U; 0 Other;
Query Match 11.5%; Score 180.8; DB 2; Length 1879;
Best Local Similarity 48.1%; Pred. No. 2.8e-22;
Matches 704; Conservative 0; Mismatches 742; Indels 18; Gaps 6;
QY 66 GCGCGGACCACTTCGACCTCGAGGAACTGCGCGCCAGCTCGTTCGGGTCGTGGC 125
DB 286 GCGAAGGATTAACGATCACTCTCGAACAACAGCCCGCTCGAGTGGCCGCGACTGAT 345
QY 126 GAGATGATGATGCTGCTTGCAGCTGATCGAGAACTACCGAAACCGCGCAGCCGCGAC 185
DB 346 GCGCGAGGACCTGGGCTCTACGAGCTGCTCGACAGTGTGACGCGGACGAGGCGACCCC 405
QY 186 GCGGTACAGGACCGCGACATCGCGCGGAGACCGGACCGGACGAGACCGTTCACGCG 245
DB 406 GCGGACCACTGACCGGAGCAGC---CGGTGCAAGTCCGAGCGGAAATCCTTTGAGCGC 462
QY 246 ATTCATCCGGTTCTCCCGGTTGAGGGCGCCACGAGGGGCACTGAGCGACCTGACCGC 305
DB 463 TTGGTATGTGACCAACAGCATCCCGCGACGTCGAGCGCGCTCTGACCGCGCGACGCT 522
QY 306 GCGGATCAAGGACTGATCGCATCGCGGTATGCCACCGACGACGCGTCCCGATGCT 365
DB 523 GCGGATCAAGGACAACTGACCGTGGCGGAGTTCGGATGATGACGAGTCTCGGACGCT 582
QY 366 CCGGACTGTGATGCGCACCGAGATGCGGTGTGTGAGGGGTGCTCGCGGCGAGGCGC 425
DB 583 AGAGGATTTACTCCGTCAAGGACGCGACTGTGTGCTACTGACTACTGCGCGCGGTGC 642
QY 426 CACCATCTCGCGACAGCACTCGAGGACATGCGGAT---GGGTATCGGTGAAGGAG 482
DB 643 AACCGTCGCGGCAAGCTGTGTGAGGACCTGTGTTCGCGGTTCCGCTTCACAC 702
QY 483 CGTCTACGCTCTGCGCTGAACCCGAAACCCCGCCACCGGACGCGGTGATCTTCAG 542
DB 703 GCGAAGCGGACCGGTCCGCAATCATGCGACCGGACGCGAGCAGTGTGATCCGG 762
QY 543 CGGCTCGCGCGTTCGCTGCGGATGTCGATGTCGCTGCGGCTGCGGCTGCGATGAGC 602
DB 763 CCGGAGTGCAGCACTCGTCGCAACCGGTGACGTCGATTTTGCATTCGCGGGGATCAAG 822
QY 603 AGCGAGCATCCGATCCCGCGCGATGTCGCGACTGTGTCGATGAAGGAGGACCGCG 662
DB 823 CGGATCGATCCGATCCCGCGCGCATTCGCGCGTGTGCGGCAACAGCCGACGTTCCG 882
QY 663 CTGCTGCGCTTTACGCGCTGACATACATGACCAACCTTGGACCAATCGGCGCCAT 722
DB 883 GCTCGTCCGATACCGGTGCAATTTCCCATCGAGCGAAACATCGACCATCTCGGCGCGAT 942
QY 723 CACGAGGGGTCGAGCTCAACCGCGGTCTCGAGGTGTTGCGCGGCGCGCTGCGC 782
DB 943 CACAGCAGGTCAGATGACGACTGATGCTCTCGGTATCGCGCGCGCGCGAGCTAA 1002
QY 783 CGACCTCAGTGGTTCGTAACCTTCGAGAGCGGAGAACTACGCTCGCGCTCGGCGA 842
DB 1003 CGACCCAC---GCCAAGCGGACAGTGTGCAAGCAGTGTACTCTGTCCACCTCGACTC 1059
QY 843 GGGATATCGGCTCTGAGATCGCGGTGCTCGAGGAGTCACTGAGCGGACGAGTGCAC 902
DB 1060 CGATGTGAGCGGCTCGAATCGGAATCGTTTCGAGAGGATTCGGGCGCGGCTCTACA 1119
QY 903 GCGGAGCTGATCGCGCGCTTCAACCGAGGACTGCGGCGCTCGAGAGCGCGGTGCGAC 962
DB 1120 GCGCGAGGTCGAGCGGAGTCCGCGACGCGCACACAGTCTGACCGGAAATCGGTTGCAC 1179
QY 963 CATCGAGCGGCTCGGTGCGGTGAGGAGCGCGGCTGCGCTATCCAGAGG---GGGT 1019
DB 1180 GGTAGAGGAAGTAAACATCCCGTGGCATCTGATGCTTTCACATCTGGAACGTGATCGC 1239

QY 1020 GATGGCTTTCAACGCGCGGCTATGCGGACTCCCGGCTGTGGGTACTTCCACAAGGG 1079
DB 1240 CACGAGCGGTGCTTACAGATGTTGGACCGCAACGATACGCGATGAACGCGGAGG 1299
QY 1080 GCGGTGAGCGTACAGCACCGCGCTCAGACGCGGCGCCAGAGTCGACACCAACAGGA 1139
DB 1300 TTGTACGATTCGGAACTGATGCGACATTTGCTTCTCGACGATTCAGACGCGGAGC 1359
QY 1140 TCTGGGATCTCTGCTCCGCGCTGATGCTGGTATCGGGAGCACCTGCGGAGCAATACCT 1199
DB 1360 TCTGTCGAAACCGTCAAACTGGTGGCCTGACCGCGCACACGCGCATCACACCTCG 1419
QY 1200 CGGATTCCTACGCGAAGGCGAGAACCTGCGGTGAGAGTTCGGCAAGCATGACAGC 1259
DB 1420 CGGCGGAGCTACGGCAAGCCGGAACCTCGTACCGCTTGCCTCCGCGCGCTACGACAC 1479
QY 1260 CGTCTCCAGGACCGGCGTGCATGCTGACCCCGCACCGCTACGTTGCCAACAGCT 1319
DB 1480 TGCCTTGACACATTCGACGCTCTGCTGATGTCGACGCTGCTGCTGCTGCTGCTG 1539
QY 1320 GTTACGCGTTCGCAAGACACCATGTCATGATCCACGATGACGGGCAATGCGATCT 1379
DB 1540 GCGCGGAGGACGATGATCGTCAACCTTCATCAC---CAAGGCTCTCGGATGATCGC 1596
QY 1380 CACACGTCGCGCTGACCTCACCGGTCAACCGGCTGACGCTGCGCCACCGGTCGCG 1439
DB 1597 CAAACGCGACCATTCGACGTCGCGGACATCCGTCCTGCTGCTGCTGCTGCTGCTG 1654
QY 1440 CGAAGAGGCTCCCGCTTGGCTTCAAGTATAGCCGCCACTTCGAGGAGTCGACGT 1499
DB 1655 -GTGAACGCGCTTCGCGTGGATGATGATCACCGGACACACTTCGACGATGCGACGT 1713
QY 1500 CTACCGCACCGCGCGCTGATCGA 1523
DB 1714 CTTCTGTCGAGCGCGCATTCGA 1737
RESULT 4
AAT62385
ID AAT62385 standard; DNA; 1566 BP.
XX AAT62385;
DT 24-JUN-1997 (first entry)
XX Rhodococcus rhodochrous amidase coding sequence.
DE Rhodococcus rhodochrous; amidase; probe; amplification; primer; PCR;
KW polymerase chain reaction; phage; open reading frame; nitrile hydratase;
XX enzyme; genetic engineering; ds.
OS Rhodococcus rhodochrous.
PH Key Location/Qualifiers
CDS 289..291
FT /*tag= a
FT /transl_except= seq: GCA, a.a.: Asp
FT CDS 316..318
FT /*tag= b
FT /transl_except= seq: ATG, a.a.: OTHER
FT /note= "OTHER designated Mec in specification"
XX JP09009973-A.
PN 14-JAN-1997.
XX 27-JUN-1995; 95JP-00184934.
XX 27-JUN-1995; 95JP-00184934.
XX (CHCC) CHISSO CORP.
XX WPI; 1997-126430/12.
DR

XX WPI: 2003-029926/02.
DR P-PSDB; ABU33813.
XX
PT New antisense nucleic acids, useful for identifying proteins or screening
PT for homologous nucleic acids required for cellular proliferation to
PT isolate candidate molecules for rational drug discovery programs.
XX
PS Claim 14; SEQ ID NO 25553; 1766pp; English.
PS
XX
CC The invention relates to an isolated nucleic acid comprising any one of
CC the 6213 antisense sequences given in the specification where expression
CC of the nucleic acid inhibits proliferation of a cell. Also included are:
CC (1) a vector comprising a promoter operably linked to the nucleic acid
CC encoding a polypeptide whose expression is inhibited by the antisense
CC nucleic acid; (2) a host cell containing the vector; (3) an isolated
CC polypeptide or its fragment whose expression is inhibited by the
CC antisense nucleic acid; (4) an antibody capable of specifically binding
CC the polypeptide; (5) producing the polypeptide; (6) inhibiting cellular
CC proliferation or the activity of a gene in an operon required for
CC proliferation; (7) identifying a compound that influences the activity of
CC the gene product or that has an activity against a biological pathway
CC required for proliferation, or that inhibits cellular proliferation; (8)
CC identifying a gene required for cellular proliferation or the biological
CC pathway in which a proliferation-required gene or its gene product lies
CC or a gene on which the test compound that inhibits proliferation of an
CC organism acts; (9) manufacturing an antibiotic; (10) profiling a
CC compound's activity; (11) a culture comprising strains in which the gene
CC product is overexpressed or underexpressed; (12) determining the extent
CC to which each of the strains is present in a culture or collection of
CC strains; or (13) identifying the target of a compound that inhibits the
CC proliferation of an organism. The antisense nucleic acids are useful for
CC identifying proteins or screening for homologous nucleic acids required
CC for cellular proliferation to isolate candidate molecules for rational
CC drug discovery programs, or for screening homologous nucleic acids
CC required for proliferation in cells other than *S. aureus*, *S. typhimurium*,
CC *K. pneumoniae* or *P. aeruginosa*. The present sequence is one of the target
CC prokaryotic essential genes. Note: The sequence data for this patent did
CC not form part of the printed specification, but was obtained in an
CC electronic format directly from WIPO at
XX ftp.wipo.int/pub/published_pat_sequences

XX SQ Sequence 1479 BP; 199 A; 542 C; 526 G; 212 T; 0 U; 0 Other;

Query Match 11.1%; Score 175.2; DB 7; Length 1479;
Best Local Similarity 49.6%; Pred. No. 2.6e-21;
Matches 636; Conservative 0; Mismatches 613; Indels 33; Gaps 6;

QY 285 GCACGTGAGCGACCTGACCGCGCGGATCAAGGACTGCATCGCCATCGCGGTATGCCAC 344
DB 213 GCCTGTGGCGGGTGGCGCTGGCGCTCAAGGACGTCTTACCACCGTCGACATGCCAC 272
QY 345 CACGACCGGTCCCGATGCTCCGACTGTGATCCGACCGAGATGCCGTGGTGGGA 404
DB 273 CACCTCGGGTCCAGATTCTCAGGGGTGGGTTGCGCGTACGACGCCACCGTCAACAC 332
QY 405 GCGGTGTCTCGCGGAGCGCCACCATCGTGGCAAGACGAACCTCGAGGACATGGCGAT 464
DB 333 GAAGCTGCGGGCGCGGATCCCGATCTCTGGGCAAGACCAACATGGACGATTCGGAT 392
QY 465 GGGTATC---GGTGAAGGAGCGTCTAGGTCCTCGGTGACCGCAACACCCCGCCCA 521
DB 393 GGGCTCTCTGACCGGAGAACTCGGCTACGCGCCGACCGCAACCCGTTGGAATGCGAGCG 452
QY 522 CGGCACGGGTGATCTTCAGCGGGTCTCGCGCGCTCGGTGCGCGCATGGTCGACTT 581
DB 453 GTTGCCCGCGGTTCGCGGGCGGAGCGCGCGCGGTGCGGGGTTCAGGCGCGCGCT 512
QY 582 CGCCCTGGCGTCTGATGAGGAGGAGGAGATCCGGATCCCGCGCGCATGGTCGGACTGGT 641
DB 513 GCGCATCGGCTCCGACACCGCGGGTTCGATCCGCGAGCGCGCGCGCTGACCGGACCGT 572
QY 642 CGGCATGAAGGAGACCCACCGGCTGGTGGCGTCTTACGGCTGACATACATGGACACAC 701

RESULT 6

AAI99683_33/c
Continuation (34 of 44) of AAI99683 from base 3300001 (Mycobacterium tuberculosis strain
WP Sequence split into 44 fragments LOCUS AAI99683 Accession AAI99683
WP Fragment Name Begin End
WP AAI99683_00 1 110000
WP AAI99683_01 100001 210000
WP AAI99683_02 200001 310000
WP AAI99683_03 300001 410000

DB 573 CGGGGTCAAGCCACCTACGACCGGTGTCGCGGTACCGGCTGGTGGCTCGCGGTCAATC 632
QY 702 CTTTGGACCATCATCGGCCCATCACAGGGGGGTTCGAGCTCAACGCCCGGGGTCTCTCAGGT 761
DB 633 GCTGGATCAGGGGGCGCGGTGGCGCGCACCGGTGTCGACACCGCGCTCTGCACCATGT 692
QY 762 GTTGGCGGGGCGACTGGCGGACCCCTCAGTGGGTGCGTAACCTTCCCGAGCGGAGAA 821
DB 693 GATCGCGGGCCACGACATCCGGGACTCCACCTCCGT---CGACGCGCGCGGTGCCACGT 749
QY 822 CTACGGCTCCCGCTCGGGGAGGAGTATCCGGTCTTGAGATTCGCGGTCTCGAGGAGTC 881
DB 750 CGTCCGCGCGCGCGCGCGCGCGCGCGCTGACCTCAAGGTGTTCGGTTCGGGTGGT 809
QY 882 ACTGGAGCGCAAGCGTGGCG-----ACGCGGACGTGATCGCGGCTTCAACAGGAGT 935
DB 810 CAAGCAGCTGCGGGGAGGAGTATCAGCCCGCGGTGCTGGGCTCTTCGAGCGCGCGGT 869
QY 936 GCGCGGGCT---CGAGAGCGCGGTGCGACCATCGAGCGGGTCTCGGTGCCGTTCGTGGAC 992
DB 870 CGAGCAGTTGACCGCGCTGGCGCGCGAGGTGAGGTGCGACTGCGCGCATCTTCGAGTA 929
QY 993 GCGGCGCTGCGCTATCCAGAGCGCGGTGATGGTTTCAAGCGCGCGGTATGGCGGACTC 1052
DB 930 CGCGCTGGCGCGCTACTACTGATTTTGGCTCGGAGGTGTCCAGCAACTTGGCCCGCTT 989
QY 1053 CGCGCGTGGGGTACTTCCAAAGGGCGCGGTGAGCTCAGCACCGCGCTCACGACGGC 1112
DB 990 CGAGCGGATGCGTACGGGCTGGCATCGGACGACGCGGACGACGAGCGGAGGT 1049
QY 1113 GGCCGAGAGTGGACCAACCCCAAGGATCTGGGATCTCTGTCGCGCTGATGCTGGTAT 1172
DB 1050 GATGGCGCTGACCGCGCGCGCGGTTCGCGCGCGGAGGTCAAAGCGCGCATCATGATCG 1109
QY 1173 CGCGGAGACCTCGCGGACGAATACCTCGGCATCCACTACGGAAGCGCGCAAGCTCGC 1232
DB 1110 CACTAGCGGTGTCGCGCGGTACTACGACGCTTACTAACACGAGGCGGAGGTGG 1169
QY 1233 GCTGGAGCTGGCAAGCAGATCGACCGCTCTTCAGGACCGGGTGTGACTGCTGACCC 1292
DB 1170 CACCTGATGCGCGCGGATCTGACCGCGCTACGAGTCCGTCGAGCTGGTGGTCTCGCC 1229
QY 1293 GACACGCTTACCGTTGCCAAGCGCTTGAGCGGTGGCAAGCACCATGTCCATGAT 1352
DB 1230 GGCACCCCGACACCGCGCTTCGGGTGGGGGAGAGGTGCGACGCCCTGGCCATGTA 1289
QY 1353 CCACCGGATGACGGGCAATCGCATCTCAACACGTGCCCGCTGGACCTCACCGGTCAACC 1412
DB 1290 TCTGTTGCA-----CCTGTGACGCTGCGCTGAACCTGGCGGGCACTG 1334
QY 1413 GCGCTGACGGTCCACCGGTG---CGGGGAGAGGGGCTGCGCGGTGGCTTGGCTTCAAGT 1469
DB 1335 CGGCATGTGGTGGCATCCGGATGTGCGCGGACGACGACGACCTGCGCGGTGGGTTCAGAT 1394
QY 1470 GATAGCCCGCATCTTCAGGAGTTCGACGCTCTACCGCACCGCGCGCGGTGATCGAGCGCG 1529
DB 1395 CATGGCGCGCGCTGCGCGGACCGGTGTGACCGGGTTCGCGCGCGCTACGAGGCCGC 1454
QY 1530 CGGCTATGGGAGCTCGCGCG 1551
DB 1455 GCGCGCGCGCTGCGGAGCGCG 1476

WP	AA199683_04	400001	510000
WP	AA199683_05	500001	610000
WP	AA199683_06	600001	710000
WP	AA199683_07	700001	810000
WP	AA199683_08	800001	910000
WP	AA199683_09	900001	1010000
WP	AA199683_10	1000001	1110000
WP	AA199683_11	1100001	1210000
WP	AA199683_12	1200001	1310000
WP	AA199683_13	1300001	1410000
WP	AA199683_14	1400001	1510000
WP	AA199683_15	1500001	1610000
WP	AA199683_16	1600001	1710000
WP	AA199683_17	1700001	1810000
WP	AA199683_18	1800001	1910000
WP	AA199683_19	1900001	2010000
WP	AA199683_20	2000001	2110000
WP	AA199683_21	2100001	2210000
WP	AA199683_22	2200001	2310000
WP	AA199683_23	2300001	2410000
WP	AA199683_24	2400001	2510000
WP	AA199683_25	2500001	2610000
WP	AA199683_26	2600001	2710000
WP	AA199683_27	2700001	2810000
WP	AA199683_28	2800001	2910000
WP	AA199683_29	2900001	3010000
WP	AA199683_30	3000001	3110000
WP	AA199683_31	3100001	3210000
WP	AA199683_32	3200001	3310000
WP	AA199683_33	3300001	3410000
WP	AA199683_34	3400001	3510000
WP	AA199683_35	3500001	3610000
WP	AA199683_36	3600001	3710000
WP	AA199683_37	3700001	3810000
WP	AA199683_38	3800001	3910000
WP	AA199683_39	3900001	4010000
WP	AA199683_40	4000001	4110000
WP	AA199683_41	4100001	4210000
WP	AA199683_42	4200001	4310000
WP	AA199683_43	4300001	4403765

Query Match 11.0%; Score 172.8; DB 4; Length 110000;

Best Local Similarity 48.9%; Pred. No. 5.6e-21;

Matches 623; Conservative 2; Mismatches 619; Indels 30; Gaps 5;

QY	275	CCACGAGGGCCACTGAGCGACCTGACCGCGCGATCAAGGACTGCATCGCCATCGCGC	334
Db	65561	CTTGGCTGGCGCTGGCGGGGTCGGCTGCGCTCAAGGACGCTTTTCAACACGAGCG	65502
QY	335	GTATGCCACACGAAACGGGTCCGGATGCTCCGACTGTGATCGCCACCGAGGATGCGC	394
Db	65501	ACATGCCACACCTCGGGTCAAAATCTTGAGGGATGCGATCTCCTACGACGCA	65442
QY	395	TGCTGTGGAGCGCTGCTCGGGGAGCGCCACCATCGTCGGCAAGACGAACTCGAGG	454
Db	65441	CGCTGACCGCGGTTCGGCGCGCGGGGATCCCGATCTCGGCAAGACCAATGAGCG	65382
QY	455	ACATGCGGATGGG---TATCGGTGAAGGACGCGTCTACGCTCTGCGTGAACCCGAA	511
Db	65381	AGTTCGGATGGCTCGTCAGCGAGAACTCCGCTTACGGTCCACCCGACACCGGTGA	65322
QY	512	ACCCGCCACGACGAGGTGATCTTCAGCGGCTCCGGCGTCCGCTGCGCGCA	571
Db	65321	ATCTACCGGTACCCGCGGTTCCGGTGGCGGACGCGCGCGCTGGCGCGGTTC	65262
QY	572	TGCTGACATTCGCGCTGGGCGTGGATGAGCGAGCATCCGGATCCCGCGCGATGTT	631
Db	65261	AGCGCGGTGGCTATCGATYCGACACCGGGGGTCTGATCCGCCAGCGCGCGGTGA	65202
QY	632	GCGGATGCTGCGATGAAGGCGACCCACGCGCTGCTGCGCTTACGCGCTGACATACA	691
Db	65201	CCGCGACCGTTCGCGCTCAAAACCCACCTACGCGACCGGTTCGCGCTATGGCTGGTGGCT	65142

QY	692	TGACACACCTTGGACCATCGGGCCCATCACCAGGGGGTGCAGCTCAACGCCCGGG	751
Db	65141	CGCGTCTCTGCTGGATCAGGGCGGCCCGTGTGCGCGACCGTCTTGACACCCCGCTGT	65082
QY	752	TCCTCGAGGTGTTGGCCGGGCGGACCTGGCGCGACCTCAGTGGGTGCTAAACCTTCGG	811
Db	65081	TGCATCAGGTGATCGCGCGGCACGACCGCGCGACTCCAGTCTCGTCCGCGAGGTGC	65022
QY	812	AGCCGGAGACTAGCGCTCCGCGCTCGCGAGGAGATCCGCTCTGAGATTCGCGGTG	871
Db	65021	CCGACGTGGTGGCGCGCTAGGCGCGCGCGGTTCGCGGATCTCTGCGGTGCGGTG	64962
QY	872	TCGAGGATCAGTGGAGCGGAAACGCTGCG-----ACGCGGAGCGTGCATCGCGCGTTCA	925
Db	64961	CGTGGTTTCGACAGCTGCACGGCGCGAGGCTACACAGCGCGGCTCTG3CCTCTTCG	64902
QY	926	ACGAGGACTGGCGCGCTCGAGAGCGCGTGGCGACCATCGAGCGGTCTCGTGGCGT	985
Db	64901	AGGCTGCGTGGAGCAGTAACCGCGTGGCGCTGAGGTGAGCGAGGTGAGTCCCGCG	64842
QY	986	TGTGACGGCGCGCTGCGCTATCCAGAGCGCGTGTGGTCTTCAACG---CGCGCGTA	1042
Db	64841	ACTTCGACCATGCTTGGCGCGCTATTACCTGATCTTGCCTCGAGGTGTCGAGCAATC	64782
QY	1043	TGGCGGACTCCGCGGTGTTGGCTACTTTCACAAGGGGCGGTGGAAGTCAAGCAACCGCG	1102
Db	64781	TGGCGCGCTTCGACGCGATGCGTACGCGGTGCGGTCGCGACGACCGCGACGCG	64722
QY	1103	TCAGACGGCGCCAGAGTCCGACACCCACAGAGTCTGGCGATCTGTCCCGGTGA	1162
Db	64721	CCGAGGAGTGTATGCGCATGACCGCGCGCGGTTCGGGCGCGAGGTCAAGCGCGCA	64662
QY	1163	TGCTGGTGTATCGGAGCACTGCGCAACGAATACCTCGGATCCACTACGCGAAGCGC	1222
Db	64661	TCATGATCGGCACTTACGCGTGTGCGCGCGCTACTACGACGCTATTACAACCGCGC	64602
QY	1223	AGAACCTGCGCTGAGCTGGCAAGCAGATCGAGCGCTCTCCAGAGCGCGCTGCAC	1282
Db	64601	AGAGGTGCGACGCTGATCCCGCGACCTCGACGCGGTATCGGTCCGTGACGTGC	64542
QY	1283	TGCTGACCCGACCACTACCTGCGCAACGAGCTGTTGAGCGGTGCGCAAGACCA	1342
Db	64541	TGGTGTGCGCCACGACCGCGCTACCGCGTTCGCGTGGGTGAGAGGTGACGATCCG	64482
QY	1343	TGTCATGATCCCAAGATGACGGGCAATGGGATCCTCAACAGTGGCGCTGACCTCA	1402
Db	64481	TGGCGATGTACTTGTTCGA-----CCTGTGACGCTGCCCTGAACCTGG	64437
QY	1403	CCGCTCACCGCGCTGACGCTGCCACCGGTG---CGGGCGAGAGGGCTGCCCGTTG	1459
Db	64436	CCGCGCACTGCGGATGTCTGTGCGCTCGGGCTCTCCCGGAGCGGGTTCGCGGTG	64377
QY	1460	GCCTCCAGTGTATAGCGCGGCACTTCAGGAGTGCAGCTCTACCGCACCGCGCGCTGA	1519
Db	64376	GCCTACAGATCATGCGCGCGGATTTGCCCGACGACCGGCTTACCGGGTGGGGCGCTT	64317
QY	1520	TCGAGGCGCGCGC 1533	
Db	64316	ATGAGGCGCGCGC 64303	

RESULT 7

ACA38250

ID ACA38250 standard; DNA; 1482 BP.

XX ACA38250;

XX AC

XX 19-JUN-2003 (first entry)

DE Prokaryotic essential gene #19907.

XX Antisense; ds; prokaryotic essential gene; cell proliferation;

XX drug design; gene.

KW

QY 1460 GCTCCAAAGTATAGCCGCCCACTTGGAGGATGCGAGTCTACCGCAGCCGCGCGTGA 1519
 Db 1388 GCCTACAGATCATGGCGCGCATTTGGCGGACGACCGGCTCTACCGGGTGGGGGCGGCTT 1447
 QY 1520 TCGAGCGCGCGCG 1533
 Db 1448 ATGAGCGCGCGCGC 1461

RESULT 8

ABZ71125
 ID ABZ71125 standard; DNA; 1482 BP.
 XX AC ABZ71125;
 XX DT 28-APR-2003 (first entry)
 XX DE Mycobacterium tuberculosis protein encoding DNA SEQ ID NO:128.
 XX KW Mycobacterium tuberculosis; infection; antibacterial; tuberculostatic;
 KW immunostimulant; vaccine; gene therapy; mycobacterial infection; gene;
 KW ds.

XX OS Mycobacterium tuberculosis.

XX PN WO2003000721-A2.

XX PD 03-JAN-2003.

XX PF 21-JUN-2002; 2002WO-GB002845.

XX PR 27-JUN-2001; 2001GB-00015365.

XX PR 07-SEP-2001; 2001GB-00021780.

XX PA (MICR-) MICROBIOLOGICAL RES AUTHORITY.

XX PI James BW, Bacon J, Marsh P;

XX P-PSDB; ABP57499.

XX WPI; 2003-201403/19.

PT New mycobacterial peptide, its fragment, variant or derivative, useful as
 PT vaccine for treating or preventing mycobacterial infections, and as
 PT diagnostic reagents for identifying such infections.

XX PS Claim 15; Page 233-234; 246pp; English.

XX CC ABP57436 to ABP57504 represent mycobacterial amino acid sequences (I)
 CC encoded by ABZ71062 to ABZ71130 (II), which are isolated from
 CC Mycobacterium tuberculosis. (I) are encoded by genes (II) whose
 CC expression is induced or up-regulated during culture of a mycobacterium
 CC under conditions defined by a dissolved oxygen tension of at least 10%
 CC air saturation measured at 37 plus degrees Celsius, when compared with a
 CC dissolved oxygen tension of at least 40% air saturation measured at 37
 CC plus degrees Celsius. (I) and (II) have antibacterial, tuberculostatic
 CC and immunostimulant activities, and can be used in vaccines and gene
 CC therapy. (I) and (II) can be used for the manufacture of a medicament for
 CC treating or preventing a mycobacterial infection. They can also be used
 CC for the manufacture of a diagnostic reagent for identifying a
 CC mycobacterial infection

XX SQ Sequence 1482 BP; 212 A; 509 C; 529 G; 232 T; 0 U; 0 Other;

Query Match 10.8%; Score 170.4; DB 7; Length 1482;
 Best Local Similarity 48.9%; Pred. No. 1.7e-20;
 Matches 623; Conservative 0; Mismatches 621; Indels 30; Gaps 5;

QY 275 CCACGGAGGGCCACTGACGACCTGACCGCGGCGATCAAGGACTGCGATGCCATCCCG 334

Db 203 CTTTGCCTCGCGCTGCGCGGGTGCCTGCGGCTCAGGACGTCTTACCACACGAGCG 262

QY 335 GTATGCCCAACCAAGAAACGGGTCCTCGGATGCTCCCGACGTGTATGCCACGAGGATGCCG 394

Db 263 ACATGCCCAACCACTGCGGGTCAAAATCTTGAGGGATGCGGATCTCCCTCATGACGCCA 322
 QY 395 TGTGTGTGAGCGGCTGCTCGCGGACAGCCCACTATGTCGCGCAAGACCACTTCGAGG 454
 Db 323 CGCTGACCGCGCGGTTGCGCGCGCGGGATCCGATCTCTGGGCAAGACCAATGAGACG 382
 QY 455 ACATGGCGATGGG---TATCGGTGAAGGAGCGTCTACGCTCTGCGCTGAACCCGGAACA 511
 Db 383 AGTTTCGGATGGGCTGCTCGACGAGAACTCCGCTTACGTTCCACCCGCAACCCGTGGA 442
 QY 512 ACCCGGCCACGCGACCGGTTGATCTTCCAGCGGCTCCGCGGCTGCGGTGCGTCCGCGCA 571
 Db 443 ATCTCGACCGGGTACCGCGGGTTCGCTGCGGACGCGCGCGCGCTGCGCGCTTCC 502
 QY 572 TGTTCGACTTGGCGCTGGCGTGCATGAGGACGAGCAGCATCCGATCCCGGCGCGCATGCT 631
 Db 503 AGGCGCGCTGGCCATCGGATCCGACACCGGGGGTTCGATCCCGCAGCCGCGCGCTGA 562
 QY 632 CGGACTGCTCGGATGAAGCGGACCCACCGGCTTGTGCGCTTTACGGGCTGACATACA 691
 Db 563 CCGGACCGTGGCGTCAAAACCCACTACGCGACGGTGTGCGCTATGCGGTGCTGCGCT 622
 QY 692 TGGACCAACCTTGGACACATCGGGGCCATCACCAGGGGGTTCGAGTCAACCGCCGCGG 751
 Db 623 CGCGTCTCTGCTGGATCAGGGCGCGCGCTGTGCGGACCGCTTTTGACACCCGCGCTGT 682
 QY 752 TCCTCGAGGTGTTGGCGGGGCGGACTGGGCGGACCCCTCACTGAGTGGGTGCTGTAACCTTCGG 811
 Db 683 TGCATCAGTGCATCGCGGCGGACGACCGCGCGGACTCCAGTCTGCTCGACCGCGAGTGC 742
 QY 812 AGCGGAGAACTACGGTTCGCGCTCGCGGAGGAGATATCCGCTCTGAGATTCGGGTCG 871
 Db 743 CCGACGTGTGGGCGCGCTAGGGCGCGCGCTGCGGGGATCTCGGTGCGGCTGCGGGTGC 802
 QY 872 TCGAGGAGTCACTGAGCGGAAACGCTGCG-----AGCGCGGAGCTGATCCCGCGGTCA 925
 Db 803 GGTGTGTCGACAGCTGCACCGCGCGGAGGCTACCGCGCGGCTGCTGCGCTCTCTTCG 862
 QY 926 ACCAGGACTGCGCGCGCTCGAGAGCGCGGTGCGGACCATCGAGCGGGTCTCGGTGCGGT 985
 Db 863 AGGCTCGCGTGGAGCAGCTAACCGCGCTGGCGCTGAGGCTCAGCGAGGTTCGACCTGCCGC 922
 QY 986 TGTGAGCGGCGCTGGCGCTATCCAGAGCGCGGTGATGGCTTTCAACG---CGCGCGCTA 1042
 Db 923 ACTTCGACCATGCCCTGCGCGCTATTACTGATCTTCTGCTCGGAGGTGTCGAGCAATC 982
 QY 1043 TGGCGGACTCCGCGCGGTGTGGGCTACTTCCAAAGGGGCGGTGGAAGTCAAGACCGCG 1102
 Db 983 TGGCGCGCTTCGACCGGATGCGCTACGGGCTGCGGGTCCGCGACGACGCGCACCCGCGCG 1042
 QY 1103 TCACAGCGGCGCGCAGAGTGCACCCACCAAGGATCTGCGGATCTGCGGCTCTGTCGCGCTGA 1162
 Db 1043 CCGAGGAGGTGATGGCGATGACCCGCGCGCGGTTTCGGGCGCGAGGTCAAGCGCGCA 1102
 QY 1163 TGCTGTGATCGCGGAGCACCTGCGCGACGAATACCTCGGATCCACTACGCGAAGGCGC 1222
 Db 1103 TCATGATCGGCACTACCGGTTGTGCGCGGCTACTACGACGCTATTACACACGAGGCGC 1162
 QY 1223 AGAACTGCGGCTGAGCTCGCGAAGAGATCGAGCGCGTCTCCAGGACCGGGTGCAC 1282
 Db 1163 AGAAGGTGCGCACGCTGATCGCGCGGCGGCTCGCGCGGCGGTATCGGTGCGTCAAGTGC 1222
 QY 1283 TGTGACCGCGGACCGGCTACCGTTGCGCAACGAGCTGTTGAGCGGTGCGGCAAGACACA 1342
 Db 1223 TGGTGTGCGGCGGACCGCGGCTTCGCGGTTGAGGAGGTGAGGATCCGATCCGCG 1282
 QY 1343 TGTTCATGATCCACGAGATGACGGGGAATGCGATCTCTMAACAGCTGCGCGCTGAGCTCA 1402
 Db 1283 TGGCGATGACTTGTTCGA-----CCTGTGCACGCTGCGCGCTGAACTTGG 1327
 QY 1403 CGGTCACCGCGCGGTGACGGTCCCGACCGG---TCCGCGCGGAGAGGCGCTTCCCGCTTG 1459

Db 1328 CCGGCACCTGCGCATGCTGTGCGCTCGGGCTCTCCCGGACGACGGTTTCGGGTTG 1387
Qy 1460 GCTTCAAGTATAGCCGCCACCTTCGAGGAGTCAGCGCTACCGCACCGCGCGCGTGA 1519
Db 1388 GCTTCAAGTATAGCGCGCGCATTTGGCGGACGACCGGCTTACCGGTGGGGCGGCTT 1447
Qy 1520 TCGAGGCGCGCGC 1533
Db 1448 ATGAGGCGCGCGC 1461

RESULT 9

ACA40772

ID ACA40772 standard; DNA; 1485 BP.

XX AC ACA40772;

XX DT 19-JUN-2003 (first entry)

XX DE Prokaryotic essential gene #22429.

XX KW Antisense; ds; prokaryotic essential gene; cell proliferation;

XX KW drug design; Gene.

XX OS Mycobacterium tuberculosis.

XX PN W0200277183-A2.

XX PD 03-OCT-2002.

XX PF 21-MAR-2002; 2002WO-US009107.

XX PR 21-MAR-2001; 2001US-00815242.

XX PR 06-SEP-2001; 2001US-00948993.

XX PR 25-OCT-2001; 2001US-0342923P.

XX PR 08-FEB-2002; 2002US-00072851.

XX PR 06-MAR-2002; 2002US-0362699P.

XX PA (ELIT-) ELITRA PHARM INC.

XX PI Wang L, Zamudio C, Malone C, Haselbeck R, Ohlsen KL, Zyskind JW;

XX PI Wall D, Trawick JD, Carr GJ, Yamamoto R, Forsyth RA, Xu HH;

XX DR WPI; 2003-029926/02.

XX DR P-PSDB; ABU36902.

XX PT New antisense nucleic acids, useful for identifying proteins or screening

XX PT for homologous nucleic acids required for cellular proliferation to

XX PT isolate candidate molecules for rational drug discovery programs.

XX PS Claim 14; SEQ ID NO 28642; 1766pp; English.

XX CC The invention relates to an isolated nucleic acid comprising any one of
XX CC the 6213 antisense sequences given in the specification where expression
XX CC of the nucleic acid inhibits proliferation of a cell. Also included are:
XX CC (1) a vector comprising a promoter operably linked to the nucleic acid
XX CC encoding a polypeptide whose expression is inhibited by the antisense
XX CC nucleic acid; (2) a host cell containing the vector; (3) an isolated
XX CC polypeptide or its fragment whose expression is inhibited by the
XX CC antisense nucleic acid; (4) an antibody capable of specifically binding
XX CC the polypeptide; (5) producing the polypeptide; (6) inhibiting cellular
XX CC proliferation or the activity of a gene in an operon required for
XX CC proliferation; (7) identifying a compound that influences the activity of
XX CC the gene product or that has an activity against a biological pathway
XX CC required for proliferation, or that inhibits cellular proliferation; (8)
XX CC identifying a gene required for cellular proliferation or the biological
XX CC pathway in which a proliferation-required gene or its gene product lies
XX CC or a gene on which the test compound that inhibits proliferation of an
XX CC organism acts; (9) manufacturing an antibiotic; (10) profiling a
XX CC compound's activity; (11) a culture comprising strains in which the gene
XX CC product is overexpressed or underexpressed; (12) determining the extent
XX CC to which each of the strains is present in a culture or collection of
XX CC strains; or (13) identifying the target of a compound that inhibits the

CC proliferation of an organism. The antisense nucleic acids are useful for
CC identifying proteins or screening for homologous nucleic acids required
CC for cellular proliferation to isolate candidate molecules for rational
CC drug discovery programs, or for screening homologous nucleic acids
CC required for proliferation in cells other than *S. aureus*, *S. typhimurium*,
CC *K. pneumoniae* or *P. aeruginosa*. The present sequence is one of the target
CC prokaryotic essential genes. Note: The sequence data for this patent did
CC not form part of the printed specification, but was obtained in
CC electronic format directly from WIPO at
CC ftp.wipo.int/pub/published_pct_sequences

XX SQ Sequence 1485 BP; 213 A; 509 C; 530 G; 233 T; 0 U; 0 Other;

Query Match 10.8%; Score 170.4; DB 7; Length 1485;

Best Local Similarity 48.9%; Pred. No. 1,7e-20;

Matches 623; Conservative 0; Mismatches 621; Indels 30; Gaps 5;

Qy 275 CCACGAGGGGCCACTGAGGACCTACCGCGGATCAAGGACTCATCGCATCGCATCGCG 334

Db 203 CCTTGGCGTGGCGCTGGCGGGTGGCGTCAAGGACGCTTCCACCACCGCG 262

Qy 335 GTATGCCACACGAGAAAGGGTCCCGGATGCTCCCGACTGTGATCGCCACCGAGGATGCCG 394

Db 263 ACATGCCACCACTCGGGTCAAAAATCTTGAGGGATGGCGATCTCCTACGACGCCA 322

Qy 395 TGGTGTGAGCGGCTGCTCGCGGAGGCGGACCATCTGCGGCAAGAGAACTCGAGG 454

Db 323 CGCTGACCGCGCGTGGCGCGCGGGATCCCGATCTCGGCAAGACCAACATGAGCG 382

Qy 455 ACATGGCGATGGG---TATCGGTGAAGGAGCGGCTCTACGGTCTCGGCTGAAACCCGAACA 511

Db 383 AGTTGGGATGGGCTCGTCAACGGAGAACTCCGCTTACGGTCCACCCGCAACCCGTGA 442

Qy 512 ACCCGGCCACGGACGGGTGGATCTTCCAGCGGCTCGCGCGTGGCGTGGCGGCA 571

Db 443 ATCTCGACCGGTACCGCGGCTTCCGCTGGCGGAGCGCGGCGGCTGGCGCGTTC 502

Qy 572 TGGTCACTTCGCGCGTGGCGTGGATGAGGAGGAGCATCCGATCCCGCGCGCATGGT 631

Db 503 AGGCGCGTGGCGATCGGATCGACACCGGGGGTGGATCCCGCAGCGCGCGCTGA 562

Qy 632 GCGGACTGTGCGCATGAAGGCGACCGACGCGCTGTGCGGCTTTACGGCTGACATACA 691

Db 563 CCGGACCGTGGCGTCAAAACCCACCTACGCGCACGGTGTGCGCTATGGGCTGGCGCT 622

Qy 692 TGGACACACATTGGGACCATCGGGCCATACACAGGGGGTTCGAGTCAACCGCGGG 751

Db 623 GCGGCTCTCGTGGATCAGGGCGGCGCGTGTGCGGCGACCGCTTGGACACCGCGTGT 682

Qy 752 TCCTCGAGGTGTTGGCGGGGCGGACTGGCGGACCGCTCAGTGGGTGCGTAACTTCGG 811

Db 683 TGCATCAGGTGATCGCGGCGCACGACCGCGCGACTCCAGTCTGGTTCGACGCGAGTGC 742

Qy 812 AGCGGAGAACTAGGCTCCGCGCTGGCGAGGAGTATCCGCTCTGAGATTCCGCGTGC 871

Db 743 CCGACGTGGTGGCGCGCGCTAGGGCGGCGCGCTGGGGGATCTGCGTGGCGTGGCGTGC 802

Qy 872 TCGAGGAGTCACTGGAGCCCAACGGTGGCG-----ACGCGGAGCGTGTATCGCGCGTCA 925

Db 803 GCGTGGTTCACAGTGCACGCGCGGAGGCGCTACCAGCGCGGCTGCTGGCTCTCTTCG 862

Qy 926 ACCAGGAGTGGCGGCGCTCGAGGCGCGGTCGACCATCGAGCGGGTCTCGGTGCGGT 985

Db 863 AGGCTGCGGTGGAGCGAGCTAACCGCGCTGGCGCGCTGAGGTCAGCGAGGTGAGTGC 922

Qy 986 TGTGAGCGGGCGCTGGCGCTTATCCAGAGCGCGGTGATGGCTTTCAAG---CGCGGGCTA 1042

Db 923 ACTTCGACCATGCGCTGGCGCGCTATTACCTGATTTGCGCTCGAGGTGTCGAGCAATC 982

Qy 1043 TGGCGGACTCGCGGTGTGGGCTACTTCCAAAGGGGCGGTGGAGTCAAGCACCGCGCG 1102

Db 983 TGGCGCGCTTCGACGCGGATCGCTACGGGCTGGCGGTCGCGGACGCGGACCGCGAGCG 1042

QY	1103	TCACGAGCGCGCCAGAGTCGACACACCACAGGATCTGGCGATCTGTCCCGGTGA	1162
Db	1043	CCGAGAGGTGATGGCGATGACCCGGCGCGGTTTCGGGCCCGAGGTCAGCGCGCA	1102
QY	1163	TGCTGGTGTATCCGGAGACCTTGGCGACGAATACCTCGGCATCCACTACGGGAAGGCGC	1222
Db	1103	TCATGATCGGCACCTACGCGTTGTGGCGGCTACTACGACGCTATTACAAACGAGGCGC	1162
QY	1223	AGAACCTGGGCTGGAGCTCGCAAGCAGATCGACGCGTCTCTCCAGAGCGGGCTGCAC	1282
Db	1163	AGAGGTGCGACGCTGATCGCCGGGACCTCGACGGCGGTATCGTCCGTGCGAGTGC	1222
QY	1283	TGCTGACCCCGACCGCTACCGTTGCGCAAGAGCTGTGACGGTGGCGAAGACACA	1342
Db	1223	TGGTGTGCGCCACGACCCCGACACCGCGTTCCGGATGGGTGAGAAGTGGACGATCCGC	1282
QY	1343	TGTCATGATCCACCGATGACGGGCAATGCGATCCTCAACAGCTGCCGCTGGACCTCA	1402
Db	1283	TGGCGATGTACTTGTTCGA-----CCTGTGCACTGTCGCGCTGAACCTGG	1327
QY	1403	CGGTCACCCCGCGCTGACGGTGCCACACGCG---TGCGGGCGAGAAGGCGCTGCCCGTTG	1459
Db	1328	CGGGCACTGGCGCATGTCTGTGCGCTCGGGCTCTCCCGGACGACGGGTTGCCGGTTG	1387
QY	1460	GCTTCCAAGTATAGCCCGCACTTCGAGAGTTCGACGCTCTACCGCACCGCGCGGTGA	1519
Db	1388	GCCTACAGATCATGGCGCGCGCATTCGCCGACGACCGGCTCTACCGGTTGGGGGGGCTT	1447
QY	1520	TCGAGGCGCGCGC 1533	
Db	1448	ATGAGGCGCGCGC 1461	

RESULT 10
AAI99682_33/c
Continuation (34 of 45) of AAI99682 from base 3300001 (Mycobacterium tuberculosis strain
WP Sequence split into 45 fragments LOCUS AAI99682 Accession Aai99682

WP	Fragment Name	Begin	End
WP	AAI99682_00	1	110000
WP	AAI99682_01	100001	210000
WP	AAI99682_02	210001	310000
WP	AAI99682_03	310001	410000
WP	AAI99682_04	410001	510000
WP	AAI99682_05	510001	610000
WP	AAI99682_06	610001	710000
WP	AAI99682_07	710001	810000
WP	AAI99682_08	810001	910000
WP	AAI99682_09	910001	1010000
WP	AAI99682_10	1010001	1110000
WP	AAI99682_11	1110001	1210000
WP	AAI99682_12	1210001	1310000
WP	AAI99682_13	1310001	1410000
WP	AAI99682_14	1410001	1510000
WP	AAI99682_15	1510001	1610000
WP	AAI99682_16	1610001	1710000
WP	AAI99682_17	1710001	1810000
WP	AAI99682_18	1810001	1910000
WP	AAI99682_19	1910001	2010000
WP	AAI99682_20	2010001	2110000
WP	AAI99682_21	2110001	2210000
WP	AAI99682_22	2210001	2310000
WP	AAI99682_23	2310001	2410000
WP	AAI99682_24	2410001	2510000
WP	AAI99682_25	2510001	2610000
WP	AAI99682_26	2610001	2710000
WP	AAI99682_27	2710001	2810000
WP	AAI99682_28	2810001	2910000
WP	AAI99682_29	2910001	3010000
WP	AAI99682_30	3010001	3110000
WP	AAI99682_31	3110001	3210000
WP	AAI99682_32	3210001	3310000
WP	AAI99682_33	3310001	3410000
WP	AAI99682_34	3410001	3510000

WP	AAI99682_35	3500001	3610000
WP	AAI99682_36	3600001	3710000
WP	AAI99682_37	3700001	3810000
WP	AAI99682_38	3800001	3910000
WP	AAI99682_39	3900001	4010000
WP	AAI99682_40	4000001	4110000
WP	AAI99682_41	4100001	4210000
WP	AAI99682_42	4200001	4310000
WP	AAI99682_43	4300001	4410000
WP	AAI99682_44	4400001	4411529

Query Match 10.8%; Score 170.4; DB 4; Length 110000;
Best Local Similarity 48.94; Pred. No. 1.5e-20; Indels 30; Gaps 5;
Matches 623; Conservative 0; Mismatches 621; Indels 30; Gaps 5;

QY	275	CCACGAGGGGCGCACTGAGCGAAGCTGACCGCCCGCGATCAAGGACTGATCGCATCGCCG 334
Db	71229	CCITTCGCTCGGCGCTGGCGGGTCCGCTCAAGGACGCTTCCACCCAGCG 71170
QY	335	GTATGCCACACGAAACGGGTCCGGATGCTCCGACTGTGATCGCCACCGAGATGCCG 394
Db	71169	ACATGCCACCCACCTGCGGGTCAAAATCTTGAGGGATGGCGATCTCCCTAGACGCCA 71110
QY	395	TGTTGGTGGAGCGGCTGCTCGGCGAGCGCCACCATCTGTCGGCAAGCAACCTCGAGG 454
Db	71109	CGCTGACCGCGGTTGGCGCGCGGGGATCCCGATCTCTGGGCAAGCAACATGACG 71050
QY	455	ACATGGCGATGGG---TATCGGTGAAGCGACGCTCTACGGTCTGCGCTGACACCGCA 511
Db	71049	AGTTTCGCGATGGGCTCGTTCGACGGAGAACTCCCGTTACGGTCCACCGCAACCCGTGA 70990
QY	512	ACCCGCGCCACGCGCGGTGATCTTCCAGCGGCTCCGGCGCTGCGGTGCTGCCCGCA 571
Db	70989	ATCTGACCGGTACCGCGGTTCCGGTGGCGGCGCGCGCGCGCTGCGCGGTTC 70930
QY	572	TGTTGACTTCCCTCGGCTCGATGAGGACGACGATCCGATCCCGCCCGATGGT 631
Db	70929	AGGCGCGCTGCCATCGATCCGACACCGGGGGTCTGATCCGCGAGCGCGCGCGTGA 70870
QY	632	GCGGACTGTCGGATGAAGCGCACCGCGCTGCTGCGCTCTTACGCGCTGCATACA 691
Db	70869	CCGCGACCGTGGCGTCAACCCACCTACGCGACGGTGTGCGCTATGGGTGTTGGGCT 70810
QY	692	TGGACACACCTTGGACACATCGGGGCCATCAACGAGGGGGTGGAGTCAACGCCCGGG 751
Db	70809	GGCGTCTCTGCTGATCAGGGCGCGCGCTGTCGGCGCACCGTCTTGGACACCGCGTGT 70750
QY	752	TCCTCGAGGTGTGGCGGGCGGACCTGCGCGGACCTCAGTGGGTGCGTAACCTTCGG 811
Db	70749	TGCATCAGGTGATCGCCCGCGCACGACCGCGCGGACTCCACGTCTGCTCGACCGCGAGTGC 70690
QY	812	AGCCCGAGAACTACCGCTCCCGCTCGCGGAGGAGTATCCGGTCTGAGATTCGCGGTGC 871
Db	70689	CCGACGTGTGGCGCGCTAGGGCGCGCGCTGCGGGATCTGCGTGGCGTTCGCGGTGC 70630
QY	872	TGAGAGGTACTGAGCGCGAACGTTGG-----ACCGCGAGGTGATCCCGCGGTCA 925
Db	70629	GCGTGTTCGACAGCTGACGCGCGGAGGCTACACGCGCGCGGTGTGCGCTCTCTTCG 70570
QY	926	ACCAGGACTGCGCGGCTTCGAGAGCGCGGTGCGACCATCGAGCGGGTCTCGGTGCGGT 985
Db	70569	AGGCTGCGGTGAGCAGCTAACCGCGTGGCGCTGAGGTGAGCGAGGTGCGACTGCCCG 70510
QY	986	TGTGACGCGCGCTGGCGCTATCCAGAGGGGGGTGATGGCTTTCAAG---CGCGGGTA 1042
Db	70509	ACTTCGACCATGCCCTGGCGCTATTACTGTTCTGCCCTCGAGGGTGTGAGCAATC 70450
QY	1043	TGGCGGACTTCCCGGTGGGCTACTTCCACAAGGGCGCGTGGACGTTCAGCACCGCGCG 1102
Db	70449	TGGCGGCTTCACGCGATGCGCTACGGGCTCGGGTGGCGACCGACCGCGAGCG 70390
QY	1103	TCACGACGCGCGCCAGAGTTCGCAACACCAAGATCTGCGGATCTCTGTCGCGGTGA 1162

DB 70389 CCGAGGAGTGATGCGATGACCCCGGCGCGCGGTTTCGGGCCCGAGGTCAAGCGGGCA 70330
QY 1163 TGCTGGTGATCGCGAGCACCTGCGCGACGAATACCTCGGCATCCACTACCGAAGGCGC 1222
DB 70329 TCATGATCGGACCTACCGGTTGTGCGCGGCTACTACGAGCGCTATTACACAGGCGC 70270
QY 1223 AGAACCTCGGGTGGAGTTCGGCAAGAGATGCGAGCGGCTCTCCAGAACCGGGTGGAC 1282
DB 70269 AGAAGGTGGCAGCTGATCGCGCGCGACCTCGACGCGCGGTATCGGTCCGTCACGTCG 70210
QY 1283 TGCTGACCCCGACCGCTACCTCGGTTGCCAAGAGCTGTGAGCGGTGCGCAAGACACCA 1342
DB 70209 TGGTGTGCGGACGACCGCGACCGCGTTCGGATGGGTGAGAGTGACATCGC 70150
QY 1343 TGTCCATGATCCACCGGATGACGGGGAATGCGATCCTCAACACGTCGCGGTGACCTCA 1402
DB 70149 TGGCGATGATCTGTTTCCA-----CCTGTGACGCTGCGGCTGAACCTGG 70105
QY 1403 CGGTCACCGCGCTGACGGTGGCCACGGG---TGCGGGGAGAGGCGCTGCCGTTG 1459
DB 70104 CCGGCACATGCGGCTGCTGTGCGGCTCTCCCGACACGCGGTTGCCGTTG 70045
QY 1460 GCTCCAAAGTATAGCGCGCCACTTCGAGGAGTCGACGCTCTACCGCACCGCGCGCTGA 1519
DB 70044 GCCTACAGATCATGCGCGCGGATTTGGCGACGACCGGCTCTACCGGTTGGGGCGGCTT 69985
QY 1520 TCGAGCGCGCGGC 1533
DB 69984 ATGAGCGCGCGGC 69971

RESULT 11

AAT62387
ID AAT62387 standard; DNA; 4730 BP.

XX AAT62387;

DT 24-JUN-1997 (first entry)

XX R. rhodochrous amidase/nitrite hydratase gene containing fragment.

DE Rhodococcus rhodochrous; amidase; probe; amplification; primer; PCR;
KW polymerase chain reaction; phage; open reading frame; nitrite hydratase;
KW enzyme; genetic engineering; ds.

XX Rhodococcus rhodochrous.

XX Key Location/Qualifiers
FH 330..334
FT RBS /*tag= a
FT /*tag= b
FT CDS /product= "amidase"
FT /note= "this fragment is missing 45 bases between
FT nucleotides 1413-1457 (nucleotides 1069-1113 of the
FT amidase gene (AAT62385))"
FT 1928..1932
FT /*tag= c
FT CDS 1939..2562
FT /*tag= d
FT /product= "nitrite hydratase alpha subunit"
FT 2589..3227
FT /*tag= e
FT /product= "nitrite hydratase beta subunit"
FT 3243..3284
FT terminator /*tag= f
FT CDS 3326..4525
FT /*tag= g
FT /product= "unknown protein"
FT /note= "this CDS is hypothetical"

PN JP09009973-A.

XX

PD 14-JAN-1997.
XX 27-JUN-1995; 95JP-00184934.
XX 27-JUN-1995; 95JP-00184934.
PR (CHCC) CHISSO CORP.
XX WPI; 1997-126430/12.
DR P-PSDB; AAW14159, AAW14160, AAW14161, AAW14162.
XX Rhodococcus rhodochrous nitrite hydratase and amidase genes - useful for
FT the industrial preparation of useful cpds.
XX Example; Page 9-15; 16pp; Japanese.
CC This is the nucleotide sequence of a fragment isolated from Rhodococcus
CC rhodochrous strain IP015564. The sequence was isolated by screening a
CC genomic R. rhodochrous DNA library in EMBL3 vector using a probe
CC amplified by the primers T62388-9. The primers amplified a 0.7 kb
CC fragment of the amidase gene. The library screen isolated a phage
CC containing a 15 kb insert. When analysed, the insert was seen to contain
CC an approx. 6 kb SacI-EcoRI fragment containing the open reading frames
CC for the amidase gene (T62385), the nitrite hydratase alpha and beta
CC subunits (T62386) and another putative open reading frame encoding a 399
CC amino acid protein (W14162) of unknown function. The enzymes, prepared by
CC genetic engineering methods, can be used in the industrial production of
CC useful compounds
XX SQ Sequence 4730 BP; 1001 A; 1517 C; 1367 G; 845 T; 0 U; 0 Other;

Query Match 10.0%; Score 157.4; DB 2; Length 4730;
Best Local Similarity 53.3%; Pred. No. 2.8e-18;
Matches 379; Conservative 0; Mismatches 326; Indels 6; Gaps 2;
QY 217 CCGACCGCGCAGAACCGGTTCAACGATTCATCCGTTTCGCGGGTGGAGGGCGCC 276
DB 534 CCAACAGCGAGCGAAATCCTTTGAGCGCTTGTGTATGTGACCAACAAGCATCCCGCGACG 593
QY 277 ACGAGGGGGCACTGAGCGACCTGACCGCGCATCAAGGACTGCTATCGCATTCGCGGT 336
DB 594 TCGAGCGGCTCTGACCGCGCGCGCATCAGGACAACTGACCGTGGCGGA 653
QY 337 ATGCCACACGACGCGGTCGCGATGCTCCCGACTGTATGCCACCGAGGATGCGGTG 396
DB 654 GTTCCGATGATGAACGGGCTCTCGGACAGTAGAGGGGTTCACTCCGCTCGCGACGGACT 713
QY 397 GTGTGGAGCGGCTGCTCGCGGCGAGCGCCACCATCGTGGCAAGACGAACTTCGAGGAC 456
DB 714 GTGATCACTCGACTACTGCGCGCGCGGTGCAACCGTTCGCGGCAAGCTGTGTGAGGAC 773
QY 457 ATGCGCAT---GGGTATCGGTGAAGCGGTCTAGCGTCTCGGTGTAACCCGAAACAC 513
DB 774 CTGTGTTTCTCCGTTTCGAGCTTCACACCGCGCAAGCGGACCGGTTCGCAATCCATGGGAC 833
QY 514 CCGCGCCACGCGACGCGGTGATCTCCAGCGGCTCGCGCGCTCCGTCGCTGCGCGCATG 573
DB 834 CCACAGCGTGAAGCAGGTGATCATCCGTTGGCAGTTCGCGCTCTGTCGCAAGCGGTGAC 893
QY 574 GTCGACTTCGCGCTGGCGCTGATGAGGAGGAGCATCCGATCCGCGCGCATGGTGC 633
DB 894 GTCGATTTTGCATCCGCGGGGATCAGGGTGGATCGATCCGATCCCGGCGCATTCGTCG 953
QY 634 GGACTGTGCGCATGAAGGCGACCCAGCGCTGGTGGCTTTACGCGCTGACATCATG 693
DB 954 GCGTGTGCGGACCAAGCGCGATTCGCGGCTCGTCCGCTATACCGGTGATTTCCCATC 1013
QY 694 GACCACACCTTGGACCATCCGCGCGCATCAGGCGGGGTGCGAGTCAACGCGCGGTC 753
DB 1014 GAGCGAACATCGACCATCTCGCGCGCATCACAGCAGCTCCACGATGCGCGACTGATG 1073
QY 754 CTCGAGGTGTGGCGGGGCGGACTGCGCGGACCCCTCAGTGGGTGGGTAACTTCGCGAG 813

Db 1074 CTCCTGGTCATCGCGGTGCGAGCGGTACGACCCAGCCAGCAGCGT---CGAA 1130
QY 814 CCGGAGAACTACGGTCCCGGTCGCGAGGAGATATCCGGTCTGAGATCCGGTCTGTC 873
Db 1131 GCAGGTGACTATCTGTCCACCTCGACTCCGATGCGATGCGATCGGATCGGT 1190
QY 874 GAGGAGTCACTGAGCGGAACGCTGCGAGCGCGGACGTCATCGCGCGTTC 924
Db 1191 CGAGAGGTTTCGGGACCGGCTTCACAGCCCGAGGTCGACGAGCGATC 1241

RESULT 12

ACA26208
ID ACA26208 standard; DNA; 1488 BP.

XX ACA26208;

DT 19-JUN-2003 (first entry)

DE Prokaryotic essential gene #7865.

KW Antisense; ds; prokaryotic essential gene; cell proliferation;
KW drug design; gene.

XX Burkholderia mallei.

XX W0200277183-A2.

XX 03-OCT-2002.

XX 21-MAR-2002; 2002WO-US0009107.

XX 21-MAR-2001; 2001US-00815242.

PR 06-SEP-2001; 2001US-00948993.

PR 25-OCT-2001; 2001US-0342923P.

PR 08-FEB-2002; 2002US-00072851.

PR 06-MAR-2002; 2002US-0362699P.

XX (BLIT-) ELITRA PHARM INC.

PI Wang L., Zamudio C, Malone C, Haselbeck R, Ohlsen KL, Zyskind JW;
PI Wall D, Trawick JD, Carr GJ, Yamamoto R, Forsyth RA, Xu HH;

DR WPI; 2003-029926/02.

DR P-PSDB; ABU22336.

XX New antisense nucleic acids, useful for identifying proteins or screening
PT for homologous nucleic acids required for cellular proliferation to
PT isolate candidate molecules for rational drug discovery programs.

XX Claim 14; SEQ ID NO 14078; 1766pp; English.

XX The invention relates to an isolated nucleic acid comprising any one of
CC the 6213 antisense sequences given in the specification where expression
CC of the nucleic acid inhibits proliferation of a cell. Also included are:
CC (1) a vector comprising a promoter operably linked to the nucleic acid
CC encoding a polypeptide whose expression is inhibited by the antisense
CC nucleic acid; (2) a host cell containing the vector; (3) an isolated
CC polypeptide or its fragment whose expression is inhibited by the
CC antisense nucleic acid; (4) an antibody capable of specifically binding
CC the polypeptide; (5) producing the polypeptide; (6) inhibiting cellular
CC proliferation or the activity of a gene in an operon required for
CC the gene product; (7) identifying a compound that influences the activity of
CC the gene product or that has an activity against a biological pathway
CC required for proliferation, or that inhibits cellular proliferation; (8)
CC identifying a gene required for cellular proliferation or the biological
CC pathway in which a proliferation-required gene or its gene product lies
CC or a gene on which the test compound that inhibits proliferation of an
CC organism acts; (9) manufacturing an antibiotic; (10) profiling a
CC compound's activity; (11) a culture comprising strains in which the gene
CC product is overexpressed or underexpressed; (12) determining the extent
CC to which each of the strains is present in a culture or collection of
CC strains; or (13) identifying the target of a compound that inhibits the

CC proliferation of an organism. The antisense nucleic acids are useful for
CC identifying proteins or screening for homologous nucleic acids required
CC for cellular proliferation to isolate candidate molecules for rational
CC drug discovery programs, or for screening homologous nucleic acids
CC required for proliferation in cells other than *S. aureus*, *S. typhimurium*,
CC *X. pneumoniae* or *P. aeruginosa*. The present sequence is one of the target
CC prokaryotic essential genes. Note: The sequence data for this patent did
CC not form part of the printed specification, but was obtained in
CC electronic format directly from WIPO at
CC ftp.wipo.int/pub/published_pct_sequences

XX Sequence 1488 BP; 234 A; 520 C; 525 G; 209 T; 0 U; 0 Other;

Query Match 9.6%; Score 152.2; DB 7; Length 1488;

Best Local Similarity 48.6%; Pred. No. 2.3e-17;

Matches 631; Conservative 0; Mismatches 623; Indels 45; Gaps 6;

QY 273 CGCCACGAGGGCGGCACTGAGCGACCTGACCGCGCATCAAGCATGTCATCGCATCGC 332

Db 183 CGGCGAGCGGGGCGCGCTCGCGGCCCTGCGCATCGTGCACAAGACGTGTGTCACGCG 242

QY 333 CGGTATGCCACACCAAGAAACGGGTCCCGATGCTCCGACTGTGATCGCCACCGAGGATGC 392

Db 243 CGGCTGGGCTCCACCGCGGCTCGAAGATGCTCGAAGTCTACGCGAGCGGCTTCGACGC 302

QY 393 CGTGTGTGTGAGCGGCTGTCGCGGAGCGGCGCCACCATCGTGGGAGAGACGACCTCGA 452

Db 303 GACCGTGTGTCGCGCGCGCTGTCGCGCGCGCGCATGCGTCAAGCATGCAATGGA 362

QY 453 GGACATGGCGATGGGT---ATCGGTGAAGCAGCGCTGTACGGTCTCTGCGCTGAACCCGAA 509

Db 363 CGATTTCGGATGGGTTCGTGCAAGAGAACTCCGCGTTCGGCCCGGTGAAGAACCGGTG 422

QY 510 CAACCCCGCCCAAGCGGCGGTGATCTTCAGCGGCTTCGGCGCTGCGTTCGTCGCGG 569

Db 423 GACACGAGCGCGGTGCGCGCGCGAGCTCGGCGCGGAGCTCCCGCGCGCTGCGCGCGG 482

QY 570 CATGTGCACTTCGCGCTGCGCGCGCGTTCGATGAGCGAGGAGCATCCGATCCCGCGCGCATG 629

Db 483 CTTGCGCGCGCGCGAGCGGCAAGACGCGGCGGTTCGATCCGCGCGCGCGGTGCTT 542

QY 630 GTGCGGACTGTGCGGATGAAAGCGAACCCAGCGCTGTCGCTTACGGCTTGACATA 689

Db 543 CGCGCGGTGACCGCGCATCAAGCGGAGCTACGCGCGGTGTCGCGCTACGCGGATGATCGC 602

QY 690 CATGACACACCTTGGACACATCGGGCCCATCACAGGGGGGTTCGAGCTCAACGCCCG 749

Db 603 GTTCGCGTGTGCTGCTGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCT 662

QY 750 GGTCTCTGAGGTGTTGGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 797

Db 663 CTTGCTGAACGCGATGGCGGCTTCGACGAGCGCGGATTCGACGAGCTTCGAGCGCGCG 722

QY 798 -----GGTAACTTCGGAGCGGAGAACTACGGCTCCGCGCTCGCGCGCGG 845

Db 723 CGAAGACTACAGCG 782

QY 846 AGTATCCGCTGTAGATTCGCGGTGTCGAGGAGTCACTGGAGCGCGAAACGGTTCGACGCC 905

Db 783 GTCGCGCGCGCTGCGCATCGCGCTTCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 839

QY 906 GGAAGTGTGCGCGCGGTTCACACGAGGAGTGGCGCGCTGAGAGCGCGCGGTGCGACCAT 965

Db 840 CGACGTGCG 899

QY 966 CGAGCGGCTCTCGGTGCGGTGTCGACGCGCGCGCTGCGCTATCCAGAGCGCGGTGATGCG 1025

Db 900 CGTTCGCGGTGTCGCTGCGCGGAGACGCGAGTGTGCTGATTCGCTACTATGTGATCGCG 959

QY 1026 TTTCACGCGCGCGCTA---TGGCGGACTCCCGCGGTGTGGCTACTTCCACAGGGGG 1082

Db 960 CGCGGAGGCGTGTGCGAACTGTGCGGTTCACGCGCGGTGCGCTACGCGCAC-----CG 1013

QY 1083 CTTGACGCTCAGACCGCGCTCAGACGGCGCGCCAGAGTCGACCCACCCACAGGATCT 1142
Db 1014 TCGGCGGATACCGGATCTCTCTGACATGTACAGAAAGTGGCGCGCGAGGCTTCGG 1073
QY 1143 GCGCATCTCTGCTCCGGCTGATCTGGTATCGCGAGACCTTGGCGGACGATACCTCGG 1202
Db 1074 GCGGAGGTGAAGCGCGGATCTCTGTCGCGACGTACGTGTCTGCGACGGCTACTACGA 1133
QY 1203 CATCACTACGGAAGGCGCAGAACTCGGCTGGAGCTCGGACGAGATCGAGCGCT 1262
Db 1134 CGCTACTACCTGAGCGCGCAGAAATCGCGGATCTCGGCGAGATTTCCAGGAAGC 1193
QY 1263 CTCCAGGACCGGGTGTACTCTGACCCCGACCGCTACCTACCTGTTGCCAACGAGCTTT 1322
Db 1194 GTTAAATCTCGACGTGATCATGGCGCGCGCTCGCCACCGTCGGTGGGACATCGG 1253
QY 1323 GAGCGTGGCAGACACCATCTCATGATCCACGATGAGGCGCATGCGATCTCTCAA 1382
Db 1254 CCGGAAGGCGACGATCTCCGTCCAGATGATC-----TGGCGGATATCTATACGCTGTC 1307
QY 1383 CACGTGCGCGCTGGACCTCACCGGTTCACCGCGCTGACGGTGCACCGGGTGGCGGGA 1442
Db 1308 GTGAGCCTCGCGGCTTCCCGGATGAGCGTGGCGCTTCGGCGGGCGGAA 1367
QY 1443 GAAGGCTCGCGTGGCTTCAAGTATAGCGCGCACTTCGAGGAGTGAAGCTCTA 1502
Db 1368 CGGGAAGCGCGGCTGGGCTGCATCATCGGCAACTATTTTCGACGAAGCGCGATGCT 1427
QY 1503 CGGACCGCGCGCTGATCGAGCGCGCGCGCTATGGGA 1541
Db 1428 GCAGGTGCGGAGCGGTTCAGCGCGGACCGACTGGCA 1466

RESULT 13

AAS54287

ID AAS54287 standard; DNA; 1455 BP.

XX

AC AAS54287;

XX

DT 13-FEB-2002 (first entry)

XX

DE Pseudomonas aeruginosa DNA for cellular proliferation protein #418.

XX

KW Antisense; ds; prokaryotic cellular proliferation gene; antibiotic;

XX

KW Antibacterial; drug design.

XX

OS Pseudomonas aeruginosa.

XX

XX WO200170955-A2.

XX

XX 27-SEP-2001.

XX

XX 21-MAR-2001; 2001WO-US009180.

XX

XX 21-MAR-2000; 2000US-0191078P.

PR

XX 23-MAY-2000; 2000US-0206948P.

PR

XX 26-MAY-2000; 2000US-0207727P.

PR

XX 23-OCT-2000; 2000US-0242578P.

PR

XX 27-NOV-2000; 2000US-0253625P.

PR

XX 22-DEC-2000; 2000US-0257931P.

PR

XX 15-FEB-2001; 2001US-0269308P.

XX

XX (BLIT-) ELITRA PHARM INC.

PA

XX

XX Haselbeck R, Ohlsen KL, Zyskind JW, Wall D, Trawick JD, Carr GJ;

PI Yamamoto RT, Xu HH;

XX

XX WPI; 2001-611495/70.

DR

XX P-FSD; AU36428.

XX

XX

PS

XX Claim 27; SEQ ID NO 7924; 511bp; English.

CC The invention relates to antisense inhibitors of genes essential to
CC prokaryotic cellular proliferation, their use in identifying the genes,
CC their use in the discovery of novel antibiotics, the essential genes
CC themselves and the encoded proteins. The prokaryotes used are *Escherichia*
CC *coli*, *Staphylococcus aureus*, *Salmonella typhi*, *Klebsiella pneumoniae*,
CC *Pseudomonas aeruginosa* and *Enterococcus faecalis*. The invention is also
CC useful for the identification of potential new targets for antibiotic
CC development. The antisense nucleic acids can also be used to identify
CC proteins used in proliferation, to express these proteins, and to obtain
CC antibodies capable of binding to the expressed proteins. The proteins can
CC be used to screen compounds in rational drug discovery programmes. The
CC antisense nucleic acid sequence is also useful to screen for homologous
CC nucleic acids which are required for cell proliferation in a wide variety
CC of organisms. The present sequence encodes an essential prokaryotic
CC cellular proliferation protein. Note: The sequence data for this patent
CC did not form part of the printed specification, but was obtained in
CC electronic format directly from WIPO at
CC ftp.wipo.int/pub/published_pct_sequences

XX SQ Sequence 1455 BP; 260 A; 543 C; 436 G; 216 T; 0 U; 0 Other;

Query Match 9.6%; Score 151; DB 4; Length 1455;

Best Local Similarity 49.0%; Pred. NO. 3.7e-17;

Matches 598; Conservative 0; Mismatches 590; Indels 33; Gaps 6;

QY 273 CGCCACGAGGGGCCACTGAGCGACCTGACCGCGCGGATCGCCACAGGACTGTCTGCACCA 332

Db 189 CGCGAGAACGGCGCTCTCGCGCGCGGATCGCCACAGGACTGTCTGCACCA 248

QY 333 CGGTATGCCACACGAAACGGGTCCCGGATGCTCCGAGCTGTGATCGCACCGAGATGC 392

Db 249 GGGCGTACGACACGAGTGGGGTTCGAAGATGCTCGAACATTCGTCTCGCCCTTACGACGC 308

QY 393 CGTGGTGTGAGCGGCTGCTCGCGGAGCGGCGCACCATCTCGCGCAAGACGACCTCGA 452

Db 309 CACGTGTGTGAGAGCTCACCGCGCGCGGTTACCTCTCGCAAGCTGAACATGA 368

QY 453 GGACATGGCGATGGG---TATCGTGAAGGACAGCTTACGCTCTGCGCTGAACCGGAA 509

Db 369 TGAATTCGCATGGGCTCGTTCGAACAGTCCAGCCACTACGGCGCGGTGAAGAACCCCTG 428

QY 510 CAACCCCGCCACGCGACGGGTGATCTTCCAGCGGCTCGGCGCTCGCTCGCTCGCGG 569

Db 429 GAGCTCTGACCGGTGCGCGGCGGCTCTCTCGCGGTTCCGCGCGGCGAGTCCGCGGCG 488

QY 570 CATGGTCACTTCGCCCTGGGCGTGTGATGAGGAGGAGGAGATCCGATCCCGCGCGCATG 629

Db 489 CTTGCTCGCGCGCGCCACCGCGCACCGATACCGCGGCTCGATCCGCCAACCGCGCGCT 548

QY 630 GTGCGGACTGTCGGCATGAAGCGACCCACCGCTTGGTGGCTCTTACGGCTTGACATA 689

Db 549 GACCAACTCACGGGATCAAGCAACCTACGCGCGGGTTCCGGTGGGCGATGATCGC 608

QY 690 CATGAGCACACCTTGGACCAATCCGGGCGCCATCACAGGGGGGTGAGTCAACGCGCG 749

Db 609 CTACGCTTCAGGCTCGACACGCGCGCGCTGGCGCGCACCGCGGAGACTCGCGCT 668

QY 750 GGTCTCTGAGGTGTGGCGGGCGGACTGCGCGGACCCCTCACTGGGTGTGTAACCTTC 809

Db 669 GATGCTGGGGGTGATGGCCCGGATTCGATCCGAAGGACTCGACCAGCT-----CGACA 722

QY 810 GGAGCCGGAGAACTACCGGCTCCCGCTCGCGGAGGAGTATCCGGTCTGAGATTCGCGGT 869

Db 723 GCGGTTGACGACTACCTGCGCCCTCTGAGAGCGCTGAGCGGCTGCGCATCGGCCT 782

QY 870 GTGAGGAGTCACTGAGCGCGAACCGTGTGAGAGCGCGGAGTGTGCGCGCTTCAACCA 929

Db 783 GCGCGGGAATATCTTCGCGCGCGCTTCGACAGCGGATCGCCAGCGGTGCTGCGCGT 842

QY 930 GGGACTGCGCGGCTCGAGAGCGCGGTGCGACCATCGAGCGGGTCTCGGTGCGGTGTG 989

Db 843 GG---TCGAGGAGCTGAAGACGCTCGCGCGCCACGCGTGAAGGACATTTCCCTGCCGGAACAT 899
QY 990 GAGCGGGGCTGCGCTATCAGAGCGCGCTGATGGCTTTCAACGCGCGCTATGCGGA 1049
Db 900 GCAGCAGCCATCCCGGCTACTACGTAATCGCGCGCGGAGCGCTCTCAACCTGTC 959
QY 1050 CTCGCGGGTGTGGCTACTTCCACAGGGGCGGTGAGCTCAGACCGCGCTCAGAC 1109
Db 960 GCCTTCGAGCGGCTGCGCTATGGCTATCGTTCGAGCGCCCGCCAGAACTGGAAGACT 1019
QY 1110 GCGCGGCCAGAGTCGACACCAACCAAGGATCTGGGATCTGCTCCCGGCTGATGCTGT 1169
Db 1020 GTACAAGCGCTCGCGCGGGAAGGCTTCGCGACGAAGTGAAGAACCGCATCATGTCTGG 1079
QY 1170 GATCGCGGACACTCGCGGACAGATACCTCGCATCCACTAGCGGAAGCGGAGAACT 1229
Db 1080 CAGCTAGCACTCTCG---GCCGGCTACTACGATGCTTATCTCGAGGCTCAGAAGAT 1136
QY 1230 GCGGCTGGAGCTCGGCAAGCAGATCGAGCGCGTCTCCAGACCGGCTGCACTGCTGAC 1289
Db 1137 TCGCGGCTGATCAAGACACTTCTGTCAGCGCTTTGCCGAAGTGAAGTATCTCTCG 1196
QY 1290 CCGGACCACTGCTACCTGTCGCAAGAGCTGTTGAGCGGTGCGCAAGACCAATGTCAT 1349
Db 1197 CCGGACCACTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1256
QY 1350 GATCCCCAGATGACGGCAATCGATCTCAACAGTGCCTGCACTGCTGCTGCTGCTGCT 1409
Db 1257 GTACCTG-----GAAGACATCTACCATCACCGCAACTCGCGGCT 1301
QY 1410 CCGGCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1469
Db 1302 GCGGGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1358
QY 1470 GATAGCGCGCACTTCAGGA 1490
Db 1359 GCTCGCGCCCTACTTCAGGA 1379

RESULT 14

ACA42618

ID ACA42618 standard; DNA; 1455 BP.

XX ACA42618;

AC ACA42618;

XX 19-JUN-2003 (first entry)

XX Prokaryotic essential gene #24275.

XX Antisense; ds; prokaryotic essential gene; cell proliferation;

XX drug design; gene.

XX Pseudomonas aeruginosa.

XX WO200277183-A2.

XX 03-OCT-2002.

XX 21-MAR-2002; 2002WO-US009107.

XX 21-MAR-2001; 2001US-00815242.

XX 06-SEP-2001; 2001US-00948993.

XX 25-OCT-2001; 2001US-0342923P.

XX 08-FEB-2002; 2002US-00072851.

XX 06-MAR-2002; 2002US-0362699P.

XX (ELIT-) ELITRA PHARM INC.

XX Wang L, Zamudio C, Malone C, Haselbeck P, Ohlsen KL, Zyskind JW;

XX Wail D, Trawick JD, Carr GJ, Yamamoto R, Forsyth RA, Xu HH;

XX WPI; 2003-029926/02.

XX P-ESDB; ABU38748.

XX

PT

PT

XX

XX

PS

XX

XX

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

XX

SQ

Query Match

Best Local Similarity 9.6%; Score 151; DB 7; Length 1455;

Matches 598; Conservative 0; Mismatches 590; Indels 33; Gaps 6;

QY 273 CGCCACGAGGGGCGACCTGAGCGACCTGACCGCGCGGATCAAGGACTGCTCGCCATCGC 332
Db 189 CGCGGAGAACGGGCGCTGCTCGCGCGCGGATCGCCACAGGACCTGTTCTGCACCA 248
QY 333 CGGTATGCCACCAAGAGGGTCCCGGATGCTCCGAGTGTGATCGCCACGAGGATGC 392
Db 249 GGGCGTACGACCAAGCTCGGTTCGAGATGCTCGAACCTCGCTCTCGCCCTACGAGC 308
QY 393 CGTGTGTGGAGCGCTGCTCGCGGAGGCGGCGGACCATCGTGGCAAGAGAACTCGA 452
Db 309 CACCGTGTGAGAGCTTCAACCGCGCGCGGCGGCTTACCTCGGCAAGCTGAATGGA 368
QY 453 GGACATGCGGATGGG---TATCGGTGAAGGAGCGGCTTACCGTCTGCTGCTGAACCCGAA 509
Db 369 TGAATTCGCGCTGCGCTCGTGAACAGTCCAGGCGGCTACGCGCGGCTGAAGAACCCCTG 428
QY 510 CAAACCCCGCCACGCGAGGGTGGATCTTCCAGCGGCTCCGCGGCTGCTGCTGCTGCGG 569
Db 429 GAGCCTCGACCGCTGCGCGGCGGCTCTCTCGCGGCGGCTTCCGCGCGGAGTGCAGCGCG 488
QY 570 CATGCTCGACTTTCGCGCTGCGGCTGCGATGAGCGAGGAGCATCCGGATCCCGCGCGCATG 629
Db 489 CTGCTCGCGCGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCT 548
QY 630 GTGCGGAGTGTGCGGATGAAGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCT 689
Db 549 GACCAACCTCACCAGGATCAAGCCAACTGACGGCGGCGGCTTCCCGCTGGGCGCATGTCG 608

QY 690 CATGGACACACTTGGACACATCGGGCCCATCCAGGGGGTGCAGCTCAAGCCCG 749
Db 609 CTACGCTTCAGCTCGACAGGCGCGCGCTGGCGGCACACGCGAGGATCGCGCT 668
QY 750 GGTCTCTCGAGGTGTGGCGGGGCCGACTGGCGGACCTCACTGGGTGCGTAACCTTCC 809
Db 669 GATGCTGGGGGTGATGGCGGATTCGATCCGAGGACTCGACAGCGT-----CGACA 722
QY 810 GGAGCCGGAGAACTAGCGCTCGCGCTCGCGAGGAGTATCGGTCTGAGATTGCGGT 869
Db 723 GCGGTGAGCACTACTGCGCGCCCTCGAGAAGCCGCTGAGCGGCTCGCATCGGCT 782
QY 870 CFTGAGGAGTACTGAGCGGAACGCTGCGACCGCGGACGTGATCGCGCGCTTCAACCA 929
Db 783 GCGCGGGAATACTTGGCGCGGCTCGACAGCCGATCGCGACGGGTCTGCGCT 842
QY 930 GGAAGTGGGGCTCGAGAGCGCGGTGGGACCATCGAGCGGTCTCGGTGCGGTG 989
Db 843 GG---TCGAGGAGCTGAAGAGCGTGGCGGCCCGGTGAAGGACATTTCCCTCCGAACAT 899
QY 990 GAGCGGGCTGCGCTTATCCAGAGCGCGTGTGTTTCAACGCGCGGCTATGCGGA 1049
Db 900 GCAGCAGCATCCCGGCTTACTAGTAATCGCGCGCGGAGGCTCTCCAACTGTC 959
QY 1050 CTCGCGGTGTGGGTACTTCCAAAGGGCGGTGGACGTGACGACCGCGGTCAAGAC 1109
Db 960 GCGCTTGCAGCGGCTGCGTATGGCTATCGTTGCGACGCGCGCGAGAACCTGGAAGACT 1019
QY 1110 GCGCGGCCAGAGTCGACCAACCCACAGGATCTGGCATCTGTCGCGGTGATGCTGT 1169
Db 1020 GTACAGGCTCGCGGCGGAAGGCTTCGCGAGCAAGTGAAGACCGCATATGTCGG 1079
QY 1170 GATCGCGGAGCACTGCGGACGAATACCTCGGCATCCACTAGCGAAGCGCAGAACCT 1229
Db 1080 CACCTACGCACTCTCG---GCCGGCTACTACGATGCTATTACTCGAGGCTCAGAAGAT 1136
QY 1230 GCGGCTGGAGCTCGGAAGCAGATCGACGCGCTCCAGGACCGGCTGCACTGCTGAC 1289
Db 1137 TCGCCGCTGATCAAGAACGACTTCTGAGCGCTTTGCGAGTGGAGTGAATCTCGG 1196
QY 1290 CCGGACACGCTTACCGTTGCCAACGAGCTGTTGAGCGGTGGCAAGACACCATGTCCAT 1349
Db 1197 CCGGACACCGCAACCCGCGCTGGAAGATCGCGGAGAGAACGACGACCCGTTTCCA 1256
QY 1350 GATCCACGATGACGGGCAATCGATCTCAACAGTGGCGCTGAGCTACCGGTCA 1409
Db 1257 GTACCTG-----GAAGACATCTACACCATCACCGCAACCTCGCGGCT 1301
QY 1410 CCGCGGCTGACGCTGCCACCGGTTGCGGCGGAGAGGCGCTGCCGTTGGCTTCAAGT 1469
Db 1302 GCGGGGCTGTCCATGCGCGCGGCTTGTGTCG---ACGGCTGCGGTGCTCAGTT 1358
QY 1470 GATAGGCGCACTTCGAGGA 1490
Db 1359 GCTCGGCGCTTACTTCAGGA 1379

RESULT 15

ABQ81844/c

ID ABQ81844 standard; DNA; 349980 BP.

XX

AC ABQ81844;

XX

DT 19-NOV-2002 (first entry)

XX

DE Bifidobacterium longum NCC2705 related nucleotide sequence SEQ ID:1100.

XX

KW Bifidobacterium longum NCC2705; Bifidobacterium; bacterial;

KW antidiarrheic; antibacterial; inhibitor of Salmonella; detection;

KW identification; lactic acid bacterium; diarrhoea; pathogenic bacteria;

KW rotavirus; food composition; pharmaceutical composition; gene; ds.

XX

OS Bifidobacterium longum.

Synthetic.

EP1227152-A1.

31-JUL-2002.

30-JAN-2001; 2001EP-00102050.

30-JAN-2001; 2001EP-00102050.

(NEST) SOC PROD NESTLE SA.

WPI; 2002-668397/72.

Novel polynucleotide comprising Bifidobacterium genome sequence useful as

a probe or primer for detecting and/or identifying Bifidobacterium longum

in a biological sample.

Disclosure; SEQ ID NO 1100; 80pp; English.

The present invention describes a polynucleotide (I) comprising a

sequence of a Bifidobacterium genome selected from the nucleotide

sequences given in ABQ81842 and ABQ81843, or a sequence exhibiting at

least 90% identity or which hybridises with the sequences given in

ABQ81842 and ABQ81843. Also described is a polynucleotide (II) encoding a

fusion protein, comprising a sequence selected from 1097 sequences given

in ABQ81842 to ABQ81843 ligated in frame to a polynucleotide encoding a

heterologous polypeptide. (I) has antidiarrheic and antibacterial

activities, and can be used as an inhibitor of Salmonella. (I) (which is

a probe) is useful for the detection and/or identification of

Bifidobacterium longum in a biological sample. A carrier containing the

lactic acid bacterium Bifidobacterium longum NCC2705 (NCIM I-2618) can be

used for preventing and/or treating diarrhoea brought about by pathogenic

bacteria and/or rotavirus. The carrier is a food composition selected

from milk, yogurt, curd, cheese, fermented milks, milk based fermented

products, ice-creams, fermented cereal based products, milk based

powders, infant formulae, pet food or a pharmaceutical composition

selected from tablets, liquid bacterial suspensions, dried oral

supplement, wet oral supplement, dry tube feeding or wet tube feeding.

(I) is useful in DNA arrays or chips to carry out analysis of the

expression of the Bifidobacterium gene. ABQ81844 to ABQ81850 represent

Bifidobacterium related nucleotide sequences given in the Sequence

Listing from the present invention but not mentioned further within the

specification. N.B. The sequence data for this patent is not represented

in the printed specification but is based on sequence information

supplied by the European Patent Office

SQ Sequence 349980 BP; 69975 A; 105045 C; 104394 G; 70566 T; 0 U; 0 Other;

Query Match 9.3%; Score 146.4; DB 6; Length 349980;

Best Local Similarity 47.0%; Pred. No. 1.8e-16;

Matches 681; Conservative 0; Mismatches 731; Indels 36; Gaps 6;

QY 100 GCCGCCCGAGCTCGTTCGGGTGCGGAGATGGTACTGCTTCCGACCTGATCGAGCA 159

Db 211290 GCCGAGCGGCGCTGCCGTCAAGAGGCGAGCTCACCTCTCGGAGCTGCTGAAGCC 211231

QY 160 CTACCGCAACCGCGGAGCGCGGACCGGCTACAGGACCGGACATCGCGCGGACCG 219

Db 211230 CACCTCAAGGTGATCGAGGCTGCGGAGCGCTCCATCAAGGCTTCTCTGAAAGTTTCGGC 211171

QY 220 ACCGCGGACGAGACCGGTTCAACGATTTCATCGGTTCTCGCGGGTGGAGGGCGGCCACG 279

Db 211170 GATGTGCTCTCGAGGAGCGGCGGCTTCGAGCCCAAGTCCCGCAAGACAGAGCGGCC 211111

QY 280 GAGGGGCGATGAGCGGACCTGACCGCGCGATCAAGGACTGATCGCCATCGCGGTATG 339

Db 211110 CTGCGCTGAGCTGGTGGTGTGCGGATCGCCATCAAGGACATGATGTCACCAAGGGGATC 211051

QY 340 CCCACCAAGAGCGGTCGCGGATCTCCGACTGTGATCGCACCGAGGATGCGGTGTG 399

Db 211050 GAAACCAAGCGCGCTTCAAGATCTCTGAAGGTGGGTGCGGCTACGAGCCACCGTC 210991

QY 400 GTGAGCGGCTGCTCGCGGAGGCGCCACCATCGTCGCAAGACGAACCTTCGAGGACATG 459
Db 210990 ATCGAAGCTCAAGGCGCGCGCATGCCGATTTCTCGCAAGACCAACCTTCGAGGATTC 210991
QY 460 GCGATGGGTATCG---GTGAAGGAGGCTGTACGGTCTCTCGGTGAACCGCAACACCC 516
Db 210930 GCCAGGCGCTCTCCACCGAGCACTCCGCTTACAGACCAACCCACACCGGTGGATAC 210871
QY 517 GCCACCGGCTGCGGTGATCTTCAGCGGCTCGCGGCTCGCGCTCGCTGCGGCTGCTGTC 576
Db 210870 GAGCGCTGCGCGGCTTCGCGGTGCTGCTCGCGCTCGCGCTCGCGCTTCGAGGCA 210811
QY 577 GACTTCGCGCTGCGGCTGCGATGAGGAGGAGAGAGATCCCGCGCGATGCTGCGGA 636
Db 210810 CCGATCGCGCTTCGCGACCGGACACTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 210751
QY 637 CTGCTGCGGCTGAAGGCGGACCCAGCGGCTGCTGCGGCTTACGGCTGACATACATGGAC 696
Db 210750 ACCGTGCGGCTCAACCCACTACCGGCGGCTCTCCGCTTCGAGGCTGCGGCTGCGG 210691
QY 697 CACACTTGAACCACTCGGCGGCTCAACAGGCGGCTCGAGTCAACGCGCGGCTGCTC 756
Db 210690 TCCTCGCTCGACAGATCGGCGGCTCTCCGCGACCGTCTCGACTCGGCTGCTGCTG 210631
QY 757 GAGGTGTTGCGCGGCGGCTGCGGCGGCTGCGGCGGCTGCGGCTGCGGCTGCGGCTG 816
Db 210630 GAGATCATCGGTGCGGCGGCTGCGGCGGCTGCGGCGGCTGCGGCTGCGGCTGCGGCT 210571
QY 817 GAGAACTACGCTCGGCGGCTGCGGCGGCTGCGGCGGCTGCGGCTGCGGCTGCGGCTG 873
Db 210570 ATGCTCGCGGCGGCTGCGGCGGCTGCGGCGGCTGCGGCGGCTGCGGCTGCGGCTG 210511
QY 874 GAGAGTCACTGAGAGCGGAGCGGCTGCGGCGGCTGCGGCGGCTGCGGCGGCTGCGG 933
Db 210510 ATCAAGGAACTCGCGGCGGCTGCGGCGGCTGCGGCGGCTGCGGCGGCTGCGGCG 210451
QY 934 CTGCGCGGCTGCGAGAGCGGCTGCGGCGGCTGCGGCGGCTGCGGCGGCTGCGGCGG 990
Db 210450 GTCGACAAGCTCAAGGAGATGGGTGCGGAGTCTGAGTCTCGGCTCGGCTCGGCT 210391
QY 991 ACGGCGGCTGCGCTTACAGAGCGGCTGAGTGGCTTCAACGGCGGCTGAGCGGCGG 1050
Db 210390 TACTCGGTAGCGCTTACTACATCATCGGCTGCGGCTGCGGCTGCGGCTGCGGCTG 210331
QY 1051 TCCGCGGCTGCGGCTTACTTCAAGGCGGCTGCGGCTGCGGCTGCGGCTGCGGCTGCG 1101
Db 210330 TACGATGCGCTGCGGCTGCGGCTGCGGCTGCGGCTGCGGCTGCGGCTGCGGCTG 210271
QY 1102 GTCAGAGCGGCGGCTGCGGCTGCGGCTGCGGCTGCGGCTGCGGCTGCGGCTGCGGCTG 1161
Db 210270 GCCAACATGATGGCTTACCCGCTGAGCGGCTTCCGCGGAGGCTGCAAGCGGCTG 210211
QY 1162 ATGCTGGTGTGCGGAGCACTGCGGAGCACTGCGGAGCACTGCGGAGCACTGCGGAG 1221
Db 210210 ATCTCGGCACTAG---CCCTGTCGCGGCTTACGAGCGCTGCTGCTGCTGCTGCT 210154
QY 1222 CAGAACTGCGGCTGCGGCTGCGGAGCAGATCGACCGGCTTCTCCAGGACCGGCTGCA 1281
Db 210153 CAGAAGGTCCGCACTGATTAAGAGACTTCAAGAGGCTTCAAGAGGCTTCAAGAGGCT 210094
QY 1282 CTGCTGAGCGGCGGCTGCGGCTGCGGCTGCGGCTGCGGCTGCGGCTGCGGCTGCGG 1341
Db 210093 CTGCTGAGCGGCGGCTGCGGCTGCGGCTGCGGCTGCGGCTGCGGCTGCGGCTGCG 210034
QY 1342 ATGCTCATGATCCAGGATGAGCGGCAATCGGATCTCTCAACAGTGCCTGCGGCTG 1401
Db 210033 CTGGCCAT-----GTACATGAACGACATCGGCACTTCCGCGCAACCTC 209989
QY 1402 ACCGTACCGGCGGCTGAGGCTGCGGCTGCGGCTGCGGCTGCGGCTGCGGCTGCGGCT 1461
Db 209988 GCGGCGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 209929
QY 1462 TCCAAAGTATAGGCGGCTTTCGAGGATGCGGCTTACCGGACCGGCGGCTGCTGATC 1521

Search completed: June 27, 2004, 05:03:01
Job time : 698.85 secs

Db 209928 TTCCAGTTTCATCGCCCGCAGCAGCGGCGGCGGCGGCGGCGGCGGCGGCTC 209869
QY 1522 GAGGCGG 1529
Db 209868 GAAGCGG 209861

GenCore version 5.1.6
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: June 27, 2004, 04:30:32 ; Search time 4525.69 Seconds

(without alignments)
10412.231 Million cell updates/sec

Title: US-10-658-691-4

Perfect score: 1578

Sequence: 1 gtgcgccaatgcgccatt.....gcgcaccggtgctgcggtag 1578

Scoring table: IDENTITY_NUC

Gapop 10.0 , Gapext 1.0

Searched: 27513289 seqs, 14931090276 residues

Total number of hits satisfying chosen parameters: 55026578

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

EST.*

1: em_estba.*

2: em_esthum.*

3: em_estin.*

4: em_estnu.*

5: em_estov.*

6: em_estpl.*

7: em_estro.*

8: em_estc.*

9: gb_est1.*

10: gb_est2.*

11: gb_hic.*

12: gb_est3.*

13: gb_est4.*

14: gb_est5.*

15: em_estfun.*

16: em_estom.*

17: em_gss_hum.*

18: em_gss_inv.*

19: em_gss_pln.*

20: em_gss_vrt.*

21: em_gss_fun.*

22: em_gss_lman.*

23: em_gss_lmus.*

24: em_gss_pro.*

25: em_gss_rod.*

26: em_gss_phg.*

27: em_gss_vrl.*

28: gb_gss1.*

29: gb_gss2.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	147	9.3	540	28	BZ893076 HL11_0118
2	145.6	9.2	600	28	BZ895409 NaRP10_01
3	131.6	8.3	583	13	BQ245721 TaE15021A
4	122.6	7.8	688	14	CB661518 OSJNEd04H

5	118.6	7.5	517	10	BZ893076
6	117	7.4	636	10	BZ893076
7	114.2	7.2	659	28	BZ891923
8	113.4	7.2	857	10	BZ893076
9	112.4	7.1	535	14	CD874694
10	111.4	7.1	640	28	AZ935507
11	107.6	6.8	604	28	AZ935188
12	106.6	6.8	2106	28	AQ012146
13	106.2	6.7	521	28	AZ934731
14	106	6.7	672	28	AZ934200
15	104.6	6.6	517	12	BG300285
16	104.2	6.6	506	14	CA685161
17	104	6.6	595	28	AZ934675
18	101.2	6.4	722	14	CB632529
19	100.2	6.3	753	14	CA239999
20	100	6.3	696	13	BZ893076
21	100	6.3	829	13	BZ893076
22	97.8	6.2	668	14	CA500482
23	96	6.1	588	14	CD208439
24	94.6	6.0	1313	28	BZ548424
25	94.2	6.0	816	14	CB659357
26	91.6	5.8	451	14	CD873667
27	91.2	5.8	515	14	CF646949
28	90.8	5.8	440	14	CB641112
29	90.6	5.7	937	14	CA289259
30	90	5.7	677	14	CA290026
31	89.6	5.7	558	10	AW564214
32	89.6	5.7	571	28	AZ934852
33	89	5.6	749	13	BQ840610
34	88.4	5.6	599	14	CA655576
35	88.2	5.6	412	28	BZ345145
36	85.8	5.4	632	14	CB851355
37	83.2	5.3	719	9	AJ612670
38	83	5.3	445	10	BE500531
39	83	5.3	742	14	CB630965
40	82.6	5.2	662	12	BM426394
41	82.2	5.2	619	9	AL504794
42	81.8	5.2	701	14	CB853233
43	81.8	5.2	773	14	CD433011
44	81	5.1	711	14	CA289750
45	78.8	5.0	814	28	BZ578429

ALIGNMENTS

RESULT 1
BZ893076
LOCUS BZ893076 540 bp DNA linear GSS 30-JUL-2003
DEFINITION HL11_0118 HL pUC18 Library Halorubrum lacusprofundi genomic 5',
genomic survey sequence.
ACCESSION BZ893076.1 GI:33343666
VERSION BZ893076.1
KEYWORDS GSS.
SOURCE Halorubrum lacusprofundi
ORGANISM Halorubrum lacusprofundi
Archaea: Euryarchaeota; Halobacteria; Halobacteriales;
Halobacteriaceae; Halorubrum.
REFERENCE 1 (bases 1 to 540)
AUTHORS Goo,Y., Roach,J., Glusman,G., Baliga,N.S., Deutsch,K., Pan,M.,
DasSarma,S., Ng,W.V. and Hood,L.
TITLE Low-pass Sequencing for Microbial Comparative Genomics
JOURNAL Unpublished (2003)
COMMENT Contact: Goo Y
Institute for Systems Biology
1441 North 34th Street, Seattle, WA 98103, USA
Tel: 206 732 1412
Fax: 206 732 1299
Email: ygoosystemsbiology.org
Seq primer: M13 Forward
Class: shotgun.
Location/Qualifiers
1..540

INSTITUTE FOR SYSTEMS BIOLOGY
1441 North 34th Street, Seattle, WA 98103, USA
Tel: 206 732 1412
Fax: 206 732 1299
Email: ygoosystemsbiology.org
Seq primer: M13 Forward
Class: shotgun.
Location/Qualifiers
1..540

FEATURES
source

```
/organism="Halorubrum lacusprofundi"
/mol_type="genomic DNA"
/strain="ATCC 49239"
/db_xref="taxon:12247"
/clone_lib="HI pUC18 Library"
/notes="Vector: pUC18; Site 1: SmaI; A shotgun library was
constructed from Halorubrum lacusprofundi genomic DNA
using pUC18/SmaI/BAP plasmid"

ORIGIN
Query Match      9.3%; Score 147; DB 28; Length 540;
Best Local Similarity 58.9%; Pred. No. 6.9e-12;
Matches 271; Conservative 0; Mismatches 186; Indels 3; Gaps 1;

QY 330 CCGCGGTATGCCACCAACGAGTCCCGGATGCTCCCGATGATCGCCACGAGGA 389
Db      |||
QY 45  CGCGGGATCCGACACCTGGGCTCCGAGATGCTCGCGACTAGTCCCGCGTACTC 105
Db      |||
QY 390 TCCCGTGGTGGAGCGCTGCTCCGCGAGCGCCACCATCTGTCGGAAGAACACT 449
Db      |||
QY 106 CGCGACTGTCGTGACCGGCTGACCGACGCGGCGCGACGGTCTCGGCAAGCAACAT 165
Db      |||
QY 450 CGAGGACATGGCGATCGG---TATCGTGAAGCAGCGTCTACGGTCTCGCTCAACCC 506
Db      |||
QY 166 GGACGAGTCTTTATGGGAACGACACCGAGACTCCGCTTCGGCCCAACGAACCC 225
Db      |||
QY 507 GAACAACCCCGCCACGCGAGCGGTGATCTTCCAGCGCTCCGCGCTGCGCTGCTGC 566
Db      |||
QY 226 TCCGACCCCGGAGCGGCTCCCGCGGCTCCTCGGCGGCTCGGCGGCGGCTCGCGC 285
Db      |||
QY 567 CGCATGTCGACTCTCGCTCGGCTGATGATGAGGAGGAGCAGCATCCCGATCCCGCGCG 626
Db      |||
QY 286 GGGCGAGGCGAGCTGGCTCCGCTCCGACAGCGGCGGTGCGTGGTTCGCGCGCGC 345
Db      |||
QY 627 ATGTCGCGAGTGGTCGGCATGAAGCGACCCACGCGCTGGTTCGCTTACGGCTGAC 686
Db      |||
QY 346 GTTCTCGGCGTCTCGGATCAAGCCACCTACGGCTCGTCTCGCGTACGGCTCGT 405
Db      |||
QY 687 ATACATGACCAACACTTCGACCATCGGCGCATCACCAGGCGGTGAGCTCAAGCC 746
Db      |||
QY 406 CGCGTACGAACTCCTCGACAGATCGGACCGATCGCGGACCGTCCGAGGCGGCGC 465
Db      |||
QY 747 CGGGTCTCGAGTGTTCGCGGCGCGAGTGGCGCGAC 786
Db      |||
QY 466 CGNCGCCCTGACGTATCGTGGAGCGGACCGCACGAC 505
Db      |||

RESULT 2
BZ895409
LOCUS
DEFINITION
BZ895409
ACCESSION
VERSION
BZ895409.1
KEYWORDS
GSS
SOURCE
Natrialba asiatica
ORGANISM
Natrialba asiatica
Archaea; Euryarchaeota; Halobacteriales; Halobacteriaceae; Natrialba.
REFERENCE
1 (bases 1 to 600)
AUTHORS
Goo V., Roach J., Glusman G., Baliga N.S., Deutsch K., Pan M.,
DasSarma S., Ng W.V. and Hood L.
JOURNAL
Low-pass Sequencing for Microbial Comparative Genomics
COMMENT
Contact: Goo V
Institute for Systems Biology
1441 North 34th Street, Seattle, WA 98103, USA
Tel: 206 732 1412
Fax: 206 732 1299
Email: ygoosystemsbiology.org
Seq primer: M13 Forward
Class: shotgun.
Location/Qualifiers

/organism="Natrialba asiatica"
/mol_type="genomic DNA"
/strain="ATCC 700177"
/db_xref="taxon:64602"
/clone_lib="Na pUC18 Library"
/notes="Vector: pUC18; Site 1: SmaI; A shotgun library was
constructed from Natrialba asiatica genomic DNA using
pUC18/SmaI/BAP plasmid"

ORIGIN
Query Match      9.2%; Score 145.6; DB 28; Length 600;
Best Local Similarity 55.7%; Pred. No. 1.1e-11;
Matches 324; Conservative 0; Mismatches 249; Indels 9; Gaps 2;

QY 407 GGCCTGCTCGCGGAGCGCCACCATCTGTCGGGAGAGAACCTCGAGGACATCGCGATGG 466
Db      |||
QY 1  GGCCTCAAAGACGCGGCGGACCATCTGTCGGCTTTTACCAACATGTTCTTTTCGGCATGG 60
Db      |||
QY 467 GT---ATCGGTGAAGGAGCGGTCTACGGTCTCGGTCTGAAACCCGAAACAAACCCGCCACG 523
Db      |||
QY 61  GTACGACACCGAGACCTCGGCTTCGGTCCGACCGACACCCCGCGCGCGGCGCACG 120
Db      |||
QY 524 GCAGCGTGTATCTTCAGCGGCTCCGCGCTCGCTCGCTCGCGGATGGTCGACTTCG 583
Db      |||
QY 121 TTCCCGCGGCTCTCTCGGCGGCTCAGCCGCGCGCTCGCGCGCGGCGGAGCGGAACTCG 180
Db      |||
QY 584 CCCTGGCGGTGATGAGGAGGAGCATCCGATCCCGCGCGCATGTCGCGGACTGTCG 643
Db      |||
QY 181 CGCTCGGCTCGATACGGGCGGTTCGATCGCTGTCCGCGCGGTTCGCGGCGTCTGTCG 240
Db      |||
QY 644 GCATGAAGGCGACCCACGCGCTCGTGGCTGCTTACGGCTTGAATACATGAGACACACCT 703
Db      |||
QY 241 GGATCAAGCGGACTTACGGCTCTGTCGGCTTACGGCTTACGGCTTACGGCTTACGGCT 300
Db      |||
QY 704 TGAACCATCTCGGCGCCATCACCAGGGGTCGAGCTCAACGCCCGGCTCTCGAGGTGT 763
Db      |||
QY 301 TAGAGAGATTCGCGCCCTTCGCGAGACCGTCGAGACGCGCGCGCGCTCTCGAGCTCA 360
Db      |||
QY 764 TGGCGGGCGAGTGGCGGACCCCTCAGTGGTGTGTAACCTTCGCGAGCGCGGAGAACT 823
Db      |||
QY 361 TCGCAGACCGGACGAGAGCGGACGCTCGAGAACGCGAACTCGAGGGCGGAACT 420
Db      |||
QY 824 ACGGCTCGGCTCGGCGAGGAGTATCGGTCTGATTCGCGGTCTCGAGGCTCTCGAGGCTC 883
Db      |||
QY 421 ACGCCGACCGCGCAACCGGCGAGCTCGACGCGCTTACGATCGGCGTCCCGGAGTGG 480
Db      |||
QY 884 TGGAGCCGAAAGTGGCGACCGCGAGCTGATCGCGCTTCAACAGGAGACTCGCGCGC 943
Db      |||
QY 481 TCGAG-----GGCGGAGCGAGCGGCTCTCGACACCTTCTGGGACGCCATCGCTGAAC 534
Db      |||
QY 944 TCGAGAGCGCGGTGGACCATCGAGGGGTCTCGGTGCGGT 985
Db      |||
QY 535 TCGAGGCCCGGCGGCGGAGTATCAGAGTCTCGCTACCGT 576
Db      |||

RESULT 3
BQ245721
LOCUS
DEFINITION
BQ245721
ACCESSION
VERSION
BQ245721.1
KEYWORDS
EST.
SOURCE
Triticum aestivum (bread wheat)
ORGANISM
Triticum aestivum
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;
Pooideae; Triticeae; Triticum.
REFERENCE
1 (bases 1 to 583)
AUTHORS
Cloutier S.
TITLE
Wheat functional genomics - Glenlea developing seeds cDNA libraries
JOURNAL
Unpublished (2002)

BQ245721
Tae15021A02R Tae15 Triticum aestivum cDNA clone Tae15021A02R, mRNA
sequence.
BQ245721
GI:20441597
Triticum aestivum (bread wheat)
Triticum aestivum
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;
Pooideae; Triticeae; Triticum.
REFERENCE
1 (bases 1 to 583)
AUTHORS
Cloutier S.
TITLE
Wheat functional genomics - Glenlea developing seeds cDNA libraries
JOURNAL
Unpublished (2002)
```

COMMENT

Contact: Dr. Sylvie Cloutier
Cereal Research Centre, Agriculture and Agri-food Canada
195 Dafoe Rd, Winnipeg, MB, Canada R3T 2M9
Tel: (204) 983-2340
Fax: (204) 983-4604
Email: scloutier@agr.gc.ca

was cloned directionally, not all sequences generated with reverse primer were from the 5' end (same with forward primer and 3' end).
Average insert size is >1.4 kb
Plate: 021 row: A column: 02
Seq primer: M13 Reverse.

FEATURES

source

1..583
/organism="Triticum aestivum"
/mol_type="mRNA"
/cultivar="Glenlea"
/db_xref="taxon:4565"
/clone="TAB15021A02R"
/tissue_type="developing seeds"
/dev_stage="15 days after anthesis"
/lab_host="E. coli DH10B"
/clone_lib="TAB15"
/note="Vector: pCMV-SPORT6.0 (Invitrogen Technologies); Site 1: NotI; Site 2: MluI; mRNA obtained from wheat seeds of cultivar Glenlea 15 days post-anthesis"

ORIGIN

Query Match 8.3%; Score 131.6; DB 13; Length 583;
Best Local Similarity 58.1%; Pred. No. 1.2e-09;
Matches 274; Conservative 0; Mismatches 189; Indels 9; Gaps 2;
QY 262 CCGGTGGAGGCGCACGAGGGGCGCATGAGCGACCTGACCGCGCGCATCAAGACTGC 321
DB 103 CGCGAGCAGCCCGCCCTGCACGCGCTCACCTTCGCCATCAAGGACATCTCGAGTCGCC 162
QY 322 ATCGCCATCGCGGTATGCCACCAAGAGCGGTCCCGGATGCTCCGACATGTATGCC 381
DB 163 GCGCGGTACCGGTTTGGACCCCGGACTGGGCGGCGAGCGACGCCCGCGCCGCC 222
QY 382 ACCGAGATGCGGTGGTGGAGCGGTGTCGCGGAGGCGCCACCATCGTGGCAAG 441
DB 223 ACGGCGCCCGCGTGTGCGCGCGTG-----GCCGCGCGCGCCACGCGGTGGGCAAG 276
QY 442 ACGACCTCGAGACATGCGATGGTAT---CGGTGAAGCAGCGTCTAGCGTCTGCG 498
DB 277 ACGGTGATGGAGATGCGCTTACAGCATCAACGCGGAGACGCGCATACGCGACCCCC 336
QY 499 CTGAACCCGAAACACCCCGCCACGCGGATGTCCTCCAGCGGTCCGCGCGTCCGCC 558
DB 337 GCGAACCCCTGCGCCCGCGCGCTCCCGCGGCTCTCCAGCGGTCCGCGTCCGCC 396
QY 559 GTGCTGCGCGCATGTGCGACTTCGCTTGGCGCTGATGAGGAGGCGAGCATCGGATC 618
DB 397 GTGCGGCGCAGCTCGCGACTTCGCTCGGACCGGACACTGGCGGCGAGGTCAGGGTG 456
QY 619 CCGGCGCATGTGGGAGTGTGCGGATGAGAGGAGACCCAGCGCTGTGTCGCTTAC 678
DB 457 CCGCGCGCATCTGCGGATCTTGGCTCCGCGCTCCAGGACTGTGTCCACCGAG 516
QY 679 GGCCTGACATATGAGCACACCTTGAACACATCGGGGCCCATCACCAGG 730
DB 517 AACGTGTGCCCATGGCGGAGATGTCGACACTGTGCGGTGTTGCTAGGG 568

RESULT 4

CB661518

LOCUS

DEFINITION CB661518 688 bp mRNA linear EST 09-APR-2003
clone OSJNE04H23.f OSJNE04H23 5', mRNA sequence.

ACCESSION

CB661518

VERSION

CB661518.1

KEYWORDS

EST.

SOURCE

Oryza sativa (japonica cultivar-group)

ORGANISM

Oryza sativa (japonica cultivar-group)

Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;
Ehrhartoideae; Oryzeae; Oryza.

REFERENCE

AUTHORS

1 (bases 1 to 688)
Jantassuriyarat,C., Lu,G., Gowda,M., Hatfield,J., Zhou,B., Mazur,B.,
Kudrna,D., Dean,R., Soderlund,C., Wing,R. and Wang,G.
Large-scale identification of ESTs involved in the interaction
between rice and Magnaporthe grisea
Unpublished (2003)

JOURNAL

COMMENT

Contact: Rod Wing

Arizona Genomics Institute

University of Arizona

Biological Sciences West, 448A, P.O. Box 210088, Tucson, AZ

85721-0088 USA

Tel: 520 626 3967

Fax: 520 621 9288

Email: http://genome.arizona.edu

PCR Primers

FORWARD: gta aaa cga cgg cca gtg

BACKWARD: gga aac agc tat gac cat g

Plate: 04 row: H column: 23

Seq primer: gta aaa cga cgg cca gtg.

Location/Qualifiers

1..688

/organism="Oryza sativa (japonica cultivar-group)"

/mol_type="mRNA"

/cultivar="Nipponbare"

/db_xref="taxon:39947"

/clone="OSJNE04H23"

/tissue_type="leaf"

/dev_stage="3 week"

/lab_host="DH10B"

/clone_lib="OSJNE"

/note="Vector: pBluescript II KS +; Site 1: EcoRI; Site 2: XhoI; 24 hrs after inoculation with Rice Blast (C9240-1)"

ORIGIN

Query Match 7.8%; Score 122.6; DB 14; Length 688;
Best Local Similarity 58.7%; Pred. No. 2.7e-08;
Matches 253; Conservative 0; Mismatches 169; Indels 9; Gaps 2;
QY 286 CCACTGAGCGACCTGACCGCGCGGATGCTCCGACTGTGATCGCCACCGAGGATGCGGTG 345
DB 96 CCCTGACAGGCTCACCCTTCTCATCAGGACATTTGACATCGCGGCGGTGAC 155
QY 346 ACG-----AACGGTCCCGATGCTCCGACTGTGATCGCCACCGAGGATGCGGTG 399
DB 156 GGGTTCGGCAACCGGACTGGGAGGACCCAGCGCGCGCGCGCGCGCGCGCGCG 215
QY 400 GTGAGCGGCTGCTCGCGGAGCGGCGCCACCATCGTGGCAAGACGACCTCGAGGACATG 459
DB 216 GTCTCGCGCGCTGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 275
QY 460 GCGATGCGGTATCGGTGAAGSCAGCG---TCTACGGTCTTCTGCTGAACCCGAAACACCCC 516
DB 276 GCATACAGCATCTATGGCGAAGACGGGCACTACGCGACGCGCGCGCGCGCGCGCG 335
QY 517 GCCCAGGCGCGGTGATCTTCAGGGCTCGGCGCTCGCGCTCGCGCTCGCGCGCGCG 576
DB 336 GGCAGAGTCCCGCGCGGATCTCCAGTGGCTCGCGCTCGCGCTCGCGCGCGCGCG 335
QY 577 GACTTCGCCCTCGGCGTTCGATGAGGAGGCGAGCATCCGATCCCGCGCGCGCGCG 636
DB 396 GACTTCTCCCTCGGACCGGACACCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 455
QY 637 CTGCTCGGCGATGAAGGCGACCGCGCGCTGTCGCTTTACGGCTGACATACATGAC 696
DB 456 ATCTACGCGCTCG 515
QY 697 CACACCTTGA 707
DB 516 CAAATGTGTGA 526

RESULT 5	BE426574	517 bp	mRNA	linear	EST 24-JUL-2000
LOCUS	WHE0336_H04_P08ZS	Wheat unstressed seedling shoot cDNA library			
DEFINITION	Triticum aestivum cDNA clone WHE0336_H04_P08				mrna sequence.
ACCESSION	B5426574				
VERSION	B5426574.1	GI:9424417			
KEYWORDS	EST				
SOURCE	Triticum aestivum (bread wheat)				
ORGANISM	Triticum aestivum				
REFERENCE	Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; Poideae; Triticeae; Triticum.				
AUTHORS	1 (bases 1 to 517)				
	Anderson,O.D., Chao,S., Choi,D.W., Close,T.J., Fenton,R.D., Han,P.S., Haia,C.C., Kang,Y.C., Lazo,G.R., Miller,R., Rausch,C.J., Seaton,C.L. and Tong,J.C.				
TITLE	The structure and function of the expressed portion of the wheat genomes				
JOURNAL	Unpublished (2000)				
COMMENT	Contact: Olin Anderson US Department of Agriculture, Agriculture Research Service, Pacific West Area, Western Regional Research Center 800 Buchanan Street, Albany, CA 94710, USA Tel: 5105595773 Fax: 5105595818 Email: andersn@w.usda.gov Sequence have been trimmed to remove vector sequence and low quality sequence with phred score less than 20 Seq primer: Stratagene SK primer. Location/Qualifiers 1..517 /organism="Triticum aestivum" /mol_type="mRNA" /cultivar="Chinese Spring" /db_xref="taxon:4565" /clone="WHE0336_H04_P08" /tissue_type="Etisolated shoot" /dev_stage="Five day old seedling" /lab_host="E. coli SOLR" /clone_lib="Wheat unstressed seedling shoot cDNA library" /note="Vector: Lambda Uni-ZAP XR, excised phagemid; Site 1: EcoRI; Site 2: XhoI; Seeds were surface-sterilized, germinated and grown aseptically in the dark at room temperature on filter paper with water, nystatin and ceftoxime in covered crystallization dishes. Shoots were harvested. The tissue, total RNA, and poly(A) RNA were prepared, a cDNA library was made, and the cDNA clones were in vivo excised to give Bluescript phagemids in the TJ Close lab (Choi, Close, Fenton) at the University of California, Riverside. Plasmid DNA preparations and DNA sequencing were performed in the OD Anderson lab (all other authors)."				
FEATURES	source				
ORIGIN					
	Query Match	7.5%	Score 118.6;	DB 10;	Length 517;
	Best Local Similarity	60.6%;	Pred. No.9.5e-08;		
	Matches 231;	Conservative	0;	Mismatches 144;	Indels 6; Gaps 2;
Qy	290	TGACGGACCTGACCGCGCGGATCAAGGACTGCATCGGCATCGCCGGTATGCCACACGGA	349		
Db	52	TGGACGGCGTCTCGTGCCTGTAAGACGAGATCGACTGTTGCCCTACCCACACAG	111		
Qy	350	ACGGGTCCCGATGCTCCGACTGTGATCGCCACCGAGATGCCGTGGTG---	406		
Db	112	CGGGACCGCGTGTCTGGGAGCGCGCGCGTGCAGGCGGACGCGCGTGCCTGCGCGC	171		
Qy	407	GGCTGCTCGGGCAGGCCACCATCTGTGGCAAGACAACTCGAGGACATGGCGATGG	466		
Db	172	ACCTGCGCGCTGGGGCGCCCTCTCTCGCGGCAAGACCAATGCAAGAGCTCGGCGCG	231		

Best Local Similarity 60.4%; Pred. No. 1.7e-07;
Matches 230; Conservative 0; Mismatches 145; Indels 6; Gaps 2;

```

QY 290 TGAGGACCTGACCGCGGATCAAGGACTGATCGCCATCGCGGTATGCCACCAAGA 349
   |||
Db 39  TGAACGGGTGTGTGGTGAAGAGAGATGACTGCTGCTTACCCACACAG 98
   |||
QY 350 ACGGTTCGGATGTCCTCGATGTCGATCGCCACGAGATGCCGTGTG---GTGAGC 406
   |||
Db 99  GCGGACGCGTGGTGTGGGAAAGGCGCGCTGCGAGCGGACGCGCGTGGTGGCG 158
   |||
QY 407 GCGTCTCGCGGACGCGGACCATCTGTCGGCAAGACGAACCTCGAGGACATGGCGATG 466
   |||
Db 159 ACCTGGGCGATCGCGGCGCTCTCGCGGCAAGACCAATGACGAGCTCGGCGCG 218
   |||
QY 467 GTATCGGTGAAGGACGGT---CTACGGTCTGCGCTGAACCCGAAACCCCGCCACG 523
   |||
Db 219 GCACGAGCGGATCAACCCGACACCGATCGGCGAGGAACCCGTACCAACGTCGCGAAG 278
   |||
QY 524 GCACGGGTGATCTCCAGCGCTCGCGGCTCGCGTCTGCTGCGGCGATGGTCTGACTTCG 583
   |||
Db 279 TGGCGGGGGCTCTCCAGCGCTCGCGGCTCGCGGCTGCTGCGCTGCGCTCTGCGCGCTCG 338
   |||
QY 584 CCTGGCGGTGATGAGGACGAGCATCCGGATCCCGGCGCATGGTCCGAGCTGGTTCG 643
   |||
Db 339 CGCTGGCTCGACGAGGAGCTCTGTCGGATGCGGCGAGCTCTGTGCGGCGGTGTGCG 398
   |||
QY 644 GCATGAGCGGACCCACGACC 664
   |||
Db 399 GCTTCAAGCCACCGCCGAC 419
   |||

```

RESULT 7
BZ891923/c 659 bp DNA linear GSS 30-JUL-2003
LOCUS Emio_0193.xl_088.ab1 Hm pUC18 Library Haloarcula marismortui
DEFINITION Genomic 5', genomic survey sequence.

ACCESSION BZ891923
VERSION BZ891923.1 GI:33342556
KEYWORDS GSS.
SOURCE Haloarcula marismortui
ORGANISM Haloarcula marismortui
Archaea; Euryarchaeota; Halobacteriales; Halobacteriaceae; Haloarcula.
REFERENCE 1 (bases 1 to 659)
AUTHORS Goo Y., Roach J., Glusman, G., Baliga, N.S., Deutsch, K., Pan, M., DasSarma, S., Ng, W.V. and Hood, L.
TITLE Low-pass Sequencing for Microbial Comparative Genomics
JOURNAL Unpublished (2003)
COMMENT Contact: Goo Y
Institute for Systems Biology
1441 North 34th Street, Seattle, WA 98103, USA
Tel: 206 732 1412
Fax: 206 732 1299
Email: ygoo@systemsbiology.org
Seq primer: M13 Forward
Class: shotgun.

FEATURES
source
1. .659
/organism="Haloarcula marismortui"
/mol_type="genomic DNA"
/strain="ATCC 43049"
/db_xref="taxon:2238"
/clone_lib="Hm pUC18 Library"
/notes="Vector: pUC18; Site_1: SmaI; A shotgun library was constructed from Haloarcula marismortui genomic DNA using pUC18/SmaI/BAP plasmid"

ORIGIN
Query Match 7.2%; Score 114.2; DB 28; Length 659;
Best Local Similarity 52.2%; Pred. No. 4.4e-07;
Matches 304; Conservative 0; Mismatches 269; Indels 9; Gaps 2;

```

QY 517 GCCACGGACACGGTGGATCTTCCACGGGTCCGGCGCTCGCTCGCGCATGGTC 576
   |||
Db 646 GGCNGCTGTCGGCGCGCTCTCTCCGGATCGGCTGCTGCTCGCGCTGCGACGCT 587
   |||
QY 577 GACTTCGCTCGGCTGCGATGAGGACGAGCATCCGATCCCGCGCATGTCGCGA 636
   |||
Db 586 GACTCGCACTCGAAGGACACTGCGGTCTATCGCTGCCCGCGCGCTTCTGTGGC 527
   |||
QY 637 CTGCTCGGATGAAGCGACCCACGCGCTTGTGCTCTTACGCGCTGACATACATGGAC 696
   |||
Db 526 GTGCTCGGATCAAGCGGACCTACGGCTTGGTTCCGGTACGGCTCTATCGGCTACGCT 467
   |||
QY 697 CACACTTGGACACATCGGCGCCATCACAGGGGGTGGAGCTCAACGCCCGGTCCTC 756
   |||
Db 466 AACAGCTCGAAGATCGGTCCCATCGCCCTCTGTTGAGGGCGCTCGGAACTGCTG 407
   |||
QY 757 GAGGTGTTGCCCGGCGCGACTGCGCGACCCCTCAGTGGGTGCGTAACCTTCGCGAGCCG 816
   |||
Db 406 GATGTCTATCGCGGCGCGACGAAACGACGCGACGACGAGGAGCGCGCGGCGGAC 347
   |||
QY 817 GAGNACTACGGCTCGCGCTCGGCGAGGAGTATCGGTCTGAGATTGCGGGTCGTCGAG 876
   |||
Db 346 GGTCTCTACGAGCGGCGCGGACGTGACGTGACGAGCTCTCTATCGCGCTTCCNACG 287
   |||
QY 877 GAGTCACTGGAGCGCAACGCTGCGACGCCGAGCTGATCGCGCGCTTCAACAGGAGCTG 936
   |||
Db 286 GACTGCTTG-----ACGGGGCTGACGAAGCGTCTGTCGAGACGTTCTCGGACGCAATG 233
   |||
QY 937 CGGCGCTCGAGAGCGCGGTGCGACCATCGAGGGGTCTCGGTGCGGTTGTGAGCGGCG 996
   |||
Db 232 GACGACCTCGAAGCCCGGCGGAGCTACCGAGGTGACCTCCCGTGGTGGAAACAC 173
   |||
QY 997 GCCTGGCTTATCCAGAGCGCGTATGGCTTTCAAACGCGCGGCTA---TGGCGGACTCC 1053
   |||
Db 172 GCGCTTACGCGGTACTACGTATCGCATCGCATCTCGAGGCGTCTCAACCTCGCGGTTTC 113
   |||
QY 1054 GCGCGTGTGGGTACTTCTCCAAAGGCGCGCGTGGAGAGTCAAGCTCAGC 1095
   |||
Db 112 GACGCGCTCGGCTACGCGCCAAATCGGCGGCTACGACGCGCAAC 71
   |||

```

RESULT 8
BZ262005 857 bp mRNA linear EST 23-OCT-2001
LOCUS HV_CEA0002M22f Hordeum vulgare seedling green leaf EST library
DEFINITION HVCDNA0004 (Blumeria challenged) Hordeum vulgare subsp. vulgare
cDNA clone HV_CEA0002M22f, mRNA sequence.
ACCESSION BZ262005
VERSION BZ262005.2 GI:13260083
KEYWORDS EST.
SOURCE Hordeum vulgare subsp. vulgare
ORGANISM Hordeum vulgare subsp. vulgare
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;
Poideae; Triticeae; Hordeum.
REFERENCE 1 (bases 1 to 857)
AUTHORS Wing, R., Close, T.J., Kleinhofs, A., Wise, R., Wei, F., Begum, D.,
Frisch, D., Yu, Y., Henry, D., Palmer, M., Rambo, T., Simmons, J.,
Choi, D.W., Fenton, R.D., Oates, R. and Main, D.
TITLE Development of a genetically and physically anchored EST resource
for barley genomics: Blumeria infected incompatible (Mla13)
seedling leaf cDNA library
JOURNAL Unpublished (2001)
COMMENT On Nov 17, 2000 this sequence version replaced gi:11192992.
Contact: Wing RA
Clemson University Genomics Institute
Clemson University
100 Jordan Hall, Clemson, SC 29634, USA
Tel: 864 656 7288
Fax: 864 656 4293
Email: rwing@clemson.edu
Total hg bases = 385
Seq primer: AATTAACCTCTCAATAAGGG

FEATURES	High quality sequence stop: 509.	
	Location/Qualifiers	
source	1. 857	
	/organism="Hordeum vulgare subsp. vulgare"	
	/mol_type="mRNA"	
	/cultivar="C116155 (M1a13)"	
	/sub_species="vulgare"	
	/db_xref="taxon:112509"	
	/clones="HV_CBA0002M2f"	
	/tissue_type="seedling green leaf"	
	/lab_host="TJC121"	
	/clone_lib="Hordeum vulgare seedling green leaf EST	
	library="HVCNDA0004 (Alumera challenged)"	
	/notes="vector: lambdaZAP; Site 1: EcoRI; Site 2: XhoI; C.I. 16155 (M1a13) plants were greenhouse grown in the R wise lab at Iowa State University, Ames, IA; 7 day old green seedlings were challenged with isolate A27 (AvrM1a13) of Blumeria graminis f. sp. hordei, and leaves were harvested 20 and 24 hr post-inoculation and snap frozen; uninoculated leaves were harvested 20 hr post-inoculation (Wei, Wise). In the TJ Close lab at the University of California, Riverside, total RNA was prepared from each sample pool, equal quantities of all three RNA pools were combined, poly(A) RNA was purified from the mixture, one cDNA library was made, and 1 million pfu were in vivo excised to give pBluescript SK(-) cDNA phagemids (Choi, Close). Phagemids were plated and picked at the Clemson University Genomics Institute (CUGI) (Begum, Palmer, Frisch, Atkins and Wing). Plasmid DNA preparations, DNA sequencing and sequence analysis were performed at CUGI (Wing, Yu, Frisch, Henry, Simmons, Oates, Rambo, Main). The sequence has been trimmed to remove vector sequence and contains a minimum of 100 bases of phred value 20 or above. For more details on library preparation and sequence analysis see http://www.genome.clemson.edu/projects/barley . To order this clone see http://www.genome.clemson.edu/orders Also see Close TJ, Wing R, Kleinhofs A, Wise R (2001) Genetically and physically anchored EST resources for barley genomics. Barley Genetics Newsletter 31:29-30. (http://wheat.pw.usda.gov/ggpages/bgn/31/cover.html)	
ORIGIN	7.2%; Score 113.4; DB 10; Length 857;	
	Query Match Best Local Similarity 55.3%; Pred. No. 6.2e-07; Matches 286; Conservative 0; Mismatches 221; Indels 10; Gaps 3;	
QY	262	CGGGTGGAGGCGGCACGAGGGCCACTGAGCGACCTGACCGCGCGATCAGGACTGC 321
DB	91	CAGGAGCGCTACCCCTGCAGGCTCACCCTTCGCCATCAGGACATCTTCTAGTCGCC 150
QY	322	ATCGCCATTCGCGGTATGCCACACGAAACGGGTCCCGATGCTCCCGACTGTGATCGCC 381
DB	151	GGCGCGGTACAGGGTTCGGACCCCGGACTGGGCGCGGACGACGCGCGCGGCCACC 210
QY	382	ACCGAGGATCCGTGGTGGTGGAGGGGTCTGTCGGGACGGCGGCACCATCTGTCGGAG 441
DB	211	ACGGCGCCCGCTGCTGGCGC-----GCTCGCGCGGCGGCGCACGGCGTGGGCATG 264
QY	442	ACGAACCTCGAGGACATGCGGATGGGTAT---CGGTGAAGGACGCTCTACGGTCTCTGCG 498
DB	265	ACGGTGTGACGAGATGGCTACAGCATCAACGGCGAGAACGCGCACTACGGCACCCCC 324
QY	499	CTGAACCGACACACCCCGCCACGGACGGGTGGATCTTCACGCGCTCGGCGCTGCC 558
DB	325	GCCAACCCCTCGGCCCGCGCGGTCCCGCGCGGTCTCTCCACGCGCTCGCGCTGCC 384
QY	559	GTGCGTGGCGGATGGTGAATTCGCCCTGGCGGTTCGATGAGGACGAGGACGATCCGGATC 618
DB	385	GTGACGCGACGCTCGCCGACTTCGCCCTCGCACCGACACCGCGGACGGTCAGGGTG 444
QY	619	CGGCGCGATGTGGG-GACTGTGCGCATGAAGCGGACCCACCGGCTGTGTGCGTCTTA 677
DB	445	CCCGCGGCATCTGCGAGCATCTTTCGGCTACGCCCTCCACCGCGAGGGCTCCGCGGA 504
QY	678	CGGCTGTGATATGACCATGACCATCTTGCAGCATCGGCGCCCATCACAGGGGGTCTGA 737
DB	505	GAACTGTCTCCCTCGGAGCAGATGTTTCGACACTCTCGGATGGTTAGCCAGAGATGTAGC 564
QY	738	GCTCAACGCCCGGGTCTCGAGGGTGTGGCGGGGCC 774
DB	565	TACATTGTCCCGTGTGAGCTAGTGTACCTGCCGCG 601
RESULT 9	CD874694	
	LOCUS	
DEFINITION	AZO3.102N04F011001 AZO3 Triticum aestivum cDNA clone AZO3102N04,	
	mRNA sequence.	
ACCESSION	CD874694	
	VERSION	
KEYWORDS	CD874694.1 GI:32558510	
	EST.	
SOURCE	Triticum aestivum (bread wheat)	
	ORGANISM	
REFERENCE	Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;	
	Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;	
AUTHORS	Poideae; Triticeae; Triticum.	
	1 (bases 1 to 535)	
TITLE	Genoplante, a major partnership french program in plant genomics	
	Unpublished (2003)	
JOURNAL	Contact: Genoplante	
	COMMENT	
FEATURES	Genoplante	
	93, rue Henri Rochefort 91025 EVRY CEDEX France	
source	Tel: 33 1 69 47 54 00	
	Fax: 33 1 69 47 54 10	
	This sequence has been generated in the framework of the french	
	plant genomics programme 'Genoplante' (http://www.genoplante.com)	
	and http://Genoplante-info.infobiogen.fr .	
	Location/Qualifiers	
	1. 535	
	/organism="Triticum aestivum"	
	/mol_type="mRNA"	
	/cultivar="recital"	
	/db_xref="taxon:4565"	
	/clone="AZO3102N04"	
	/tissue_type="leaf"	
	/clone_lib="AZO3"	
ORIGIN	7.1%; Score 112.4; DB 14; Length 535;	
	Query Match Best Local Similarity 60.1%; Pred. No. 7.7e-07; Matches 226; Conservative 0; Mismatches 141; Indels 9; Gaps 2;	
QY	279	GGAGGGGCGACCTGAGCGACCTGACCGCGCGATCAAGGACTGCATCGCATCGCGGTAT 338
DB	138	GCAGCTACCCCTGCACGGCTCACCTTCGCTATCAAGGACATCTTCGACGTCGCGGCGC 197
QY	339	GCCCACCGACGACGGTCCC-----GGATGTCCTCCGACTGATCGCCACCGAGATGC 392
DB	198	CGTACCGGCTTCGCGACCCCGGACTGGGCCCGGACGACGCGCGCGCGCGCGCAACCGC 257
QY	393	CGTGTGTGTGAGCGGCTGCTCGCGGCGAGCGCCACCATCTCGTCGGCAAGACGAACTCGA 452
DB	258	CCCCCGCTTCCTGGCGCGCTCGCCCGCGCCACAGCGGTGGGCAAGAGCGTGTATGA 317
QY	453	GGACATGGCGATGGGTAT---CGGTGAAGGACGCTTACGGTCTCTGCTGTAACCCGAA 509
DB	318	CGAGATGGGCTTACAGCATCAACGCGGAGAACGCGCACTACGGCACCCCGCGCAACCCCTG 377
QY	510	CAACCCCGCCACCGCACGGGTGGATCTTCCAGCGGCTCCGGCGCTGCGCTGCTGCGCGG 569
DB	378	CGCCCTTGGCGGCTCCCGCGGCTCTCTCCAGCGGCTCCGCGCTCGCGCTCGCGGCCAG 437
QY	570	CATGTGCGATCTTGGCCCTTGGGCGTGTGATGAGCGACGCGAGCATCCCGGCGCGCATG 629

```
Db 438 CCTCGCCGACTTCGCCCTCGGACCGACCGGCGGACGCTCAGGTTGCCCGCGCCTA 497
QY 630 GTGCGGACTGTCGCGC 645
Db 498 CTGCGGCATCTTCGCGC 513

RESULT 10
AZ935507/c
LOCUS AZ935507 640 bp DNA linear GSS 24-APR-2001
DEFINITION BJ_Ba0003L22f B. japonicum BAC library Bradyrhizobium japonicum
genomic, genomic survey sequence.
ACCESSION AZ935507
VERSION AZ935507.1 GI:13778331
KEYWORDS GSS
SOURCE Bradyrhizobium japonicum
ORGANISM Bradyrhizobium japonicum
Bacteria; Proteobacteria; Alphaproteobacteria; Rhizobiales;
Bradyrhizobiaceae; Bradyrhizobium.
REFERENCE 1 (bases 1 to 640)
AUTHORS Tomkins,J.P., Wood,T.C., Stacey,M.G., Loh,J.T., Judd,A.,
Goicoechea,J.L., Stacey,G., Sadowsky,M.J. and Wing,R.A.
A marker-dense, sequence-ready map of the Bradyrhizobium japonicum
Genome
JOURNAL Genome Res. 11 (8), 1434-1440 (2001)
MEDLINE 21376150
PUBMED 11483585
COMMENT Contact: Wing RA
Clemson University Genomics Institute
Clemson University
100 Jordan Hall, Clemson, SC 29634, USA
Tel: 864 656 7288
Fax: 864 656 4293
Email: rwing@clemson.edu
Class: BAC ends
High quality sequence stop: 582.
FEATURES
source
1..840
/organism="Bradyrhizobium japonicum"
/mol_type="genomic DNA"
/strain="USD110"
/db_xref="taxon:375"
/lab_host="E. coli"
/clone_lib="B. japonicum BAC library"
/notes="vector: pindigo536; Site_1: HindIII"

ORIGIN
Query Match 7.1%; Score 111.4; DB 28; Length 640;
Best Local Similarity 58.3%; Pred. No. 1.1e-06;
Matches 215; Conservative 0; Mismatches 151; Indels 3; Gaps 1;

QY 257 TCTGCGGTTGAGGGCGCCACGAGGGCCACTGAGCGACTGACCGCGCGATCAAGG 316
Db 369 TCTGCGATGCGGCGCGCGCGGATCGGACCGCTGATGCGGTGCGGATCAAGG 310

QY 317 ACTGCAATCGCATCGCGGTATCCACACGAGCGGTCGCCGATGCTCCCGACTGTGA 376
Db 309 ACCTACCGATACCGGGGCTGACCACTACCGCTCGGCTTGTTCGCGATCAGC 250

QY 377 TCGCCACCGAGGATGCGGTGTTGGAGCGGTGCTCGCGGAGCGGCCACCATCGTGC 436
Db 249 TCCCGGCCGAGGACGAGCTGTTGTCGCGCGGTGCGGCGCGCGCGGATCATTTGG 190

QY 437 GCAGAGCACTTCGAGGACATGCGATGGGT---ATCGGTGAGGACCGCTACGCTC 493
Db 189 GAAAGACCAACACCGCGAATTCGGTTTCGGTCCGCTCTGCACCAACCGTTGCGCGGC 130

QY 494 CTGCGTGAACCCGAAACACCCGCCACCGCACTGATCTTCAGCGGCTCCGGCG 553
Db 129 CGACGGCAATCCCTTCGATCCCGCGCTGACCTCCGGCGGATGTCGCGCGGCTCTGCGG 70

QY 554 CTGCGTCTGTCGCGCATGTCGATTCGCCCTTGGCGGTGATGAGGACGAGCATCC 613
```

```
Db 69 TTGCGGTGCGGCGCGCATGTTAACGCTGGCGACCGACCGATTTCGGGGCTCGGTGC 10
QY 614 GGATCCCGG 622
Db 9 GCACGCCCG 1

RESULT 11
AZ935188
LOCUS AZ935188 604 bp DNA linear GSS 24-APR-2001
DEFINITION BJ_Ba0003B03f B. japonicum BAC library Bradyrhizobium japonicum
genomic, genomic survey sequence.
ACCESSION AZ935188
VERSION AZ935188.1 GI:13777512
KEYWORDS GSS
SOURCE Bradyrhizobium japonicum
ORGANISM Bradyrhizobium japonicum
Bacteria; Proteobacteria; Alphaproteobacteria; Rhizobiales;
Bradyrhizobiaceae; Bradyrhizobium.
REFERENCE 1 (bases 1 to 604)
AUTHORS Tomkins,J.P., Wood,T.C., Stacey,M.G., Loh,J.T., Judd,A.,
Goicoechea,J.L., Stacey,G., Sadowsky,M.J. and Wing,R.A.
A marker-dense, sequence-ready map of the Bradyrhizobium japonicum
Genome
JOURNAL Genome Res. 11 (8), 1434-1440 (2001)
MEDLINE 21376150
PUBMED 11483585
COMMENT Contact: Wing RA
Clemson University Genomics Institute
Clemson University
100 Jordan Hall, Clemson, SC 29634, USA
Tel: 864 656 7288
Fax: 864 656 4293
Email: rwing@clemson.edu
Class: BAC ends
High quality sequence stop: 553.
FEATURES
source
1..604
/organism="Bradyrhizobium japonicum"
/mol_type="genomic DNA"
/strain="USD110"
/db_xref="taxon:375"
/lab_host="E. coli"
/clone_lib="B. japonicum BAC library"
/notes="vector: pindigo536; Site_1: HindIII"

ORIGIN
Query Match 6.8%; Score 107.6; DB 28; Length 604;
Best Local Similarity 53.1%; Pred. No. 4e-06;
Matches 277; Conservative 0; Mismatches 239; Indels 6; Gaps 2;

QY 270 GGGCGCCACGAGGGGCGCCACTGAGCGACTGACCGCGCGATCAAGGACTGCATCGCAT 329
Db 11 GGATGCGCGCAGCATGCGGCTCTATGCGGTCCCGGTGAGGACAAACATCGACGC 70

QY 330 CGCGGTATGCCACCAACGAGCGGTCCCGAGTGTCTCCGACTGTGATCGCCACCGAGGA 389
Db 71 GCTGGGCTTTCGACACG---GCGGCTTCCCGCGCTTCTCTACACCGACCCATGA 127

QY 390 TCCCGTGGTGTGAGCGGCTGCTCGCGGAGGGGCCACCATGCTCGGAGAGACACT 449
Db 128 CTCGACCGCGGTGAGCGGCTGCGCGGCGCGCGCATCATCATCGGCAAGACCAATCT 187

QY 450 CGAGGACATGCGCATGGG---TATCGGTGAAGGCGAGGTCTACGGTCTCGGCTGAACCC 506
Db 188 CGACCATGTCGCAACCGGCGCTGTCGCGGTACGTCGCCCTACGGCATTCGCAAAATTC 247

QY 507 GAACAACCCGCCCCACCGGACGGGTGATCTTCAGCGGTTCGCGCGCTGCGCTGCTGTCG 566
Db 248 GATTTCGCGAGGATCTGATTCGCGGTGCTTCAGTTCGCGGTTCGCGGACCGCCCTCGGCGC 307

QY 567 CGGCATGTCGACTTCGCCCTTGGCGGTGATGAGGAGGAGCATCCGGATCCCGGCGGC 626
```

```

Db 308 CGGGTGTGCGCTGACGCTCGGACCGACACCGCGCGGCGCGCTGCGGCGCAT 367
Qy 627 ATGTGCGGACTGTGCGCATGAGGCGACCCAGCGCTGTGCGCTTACGCGCTGAC 686
Db 368 GCTCAACAACATCTGCGGCTGAGACCAAGCTCGGCATGATCTCGACCGCGGACTCGT 427
Qy 687 ATACATGAGACACACCTTGACACATCTCGGCCCATCACCAGGGGGTTCGAGTCAACGC 746
Db 428 GCCTGCTGCGGACGCTGACTCATCTCGTATTGCTTACAGGTGACGATGCTGC 487
Qy 747 CCGGGTCTCGAGTGTGCGGGGCGGACTGCGGCGGACCC 788
Db 488 ACTGCGACTTTCGTGATGCGGGACCTGACCAAGCGGATCC 529

RESULT 12
LOCUS AQ012146/c
DEFINITION 875011B093011698 Cosmid library of chromosome II Rhodobacter
sphaeroides genomic clone 875011B093011698, genomic survey
sequence.
ACCESSION AQ012146
VERSION AQ012146.1 GI:3177101
KEYWORDS GSS.
SOURCE Rhodobacter sphaeroides
ORGANISM Rhodobacter sphaeroides
Bacteria; Proteobacteria; Alphaproteobacteria; Rhodobacteriales;
Rhodobacteraceae; Rhodobacter.
REFERENCE 1 (bases 1 to 2106)
AUTHORS Choudhary,M., Mackenzie,C., Mouncey,N., Weinstein,G.M. and
Kaplan,S.
TITLE RsgDB, the Rhodobacter sphaeroides Genome Database
JOURNAL Unpublished (1998)
COMMENT Contact: Choudhary, M.
Department of Microbiology and Molecular Genetics
University of Texas Medical School
6431 Fannin Street, Houston, TX 77030, USA
Tel: 713 500 5437
Fax: 713 500 5499
Email: madhu@utmsi.med.uth.tmc.edu
Seq primer: pBluescript T3
Class: shotgun.

FEATURES
source
1. .2106
/organism="Rhodobacter sphaeroides"
/mol_type="genomic DNA"
/strain="2.4.1T"
/db_xref="taxon:1063"
/clone="875011B093011698"
/lab_hosts="E. coli S17-1"
/clone_lib="Cosmid library of chromosome II"
/note="Vector: pLA2917"

ORIGIN
Query Match 6.8%; Score 106.6; DB 28; Length 2106;
Best Local Similarity 52.3%; Pred. No. 7.6e-06;
Matches 285; Conservative 0; Mismatches 254; Indels 6; Gaps 2;

Qy 295 GACTGACCGCGCGGATCAGGACTGATCGCCATCGCGGTATGCCACACGACGCGG 354
Db 895 GGATCTCTGCGCGCTGAGGACGAGACCATGATGCGGGCAGGTTCAGACTACGCG 836
Qy 355 TCCGGATGTCGCCGACTGTGATGCCACCGAGATGCGGTGTTGGAGCGGCTGCTC 414
Db 835 TCGTCTTTTACGCGGACCATGTGCTGCGAAGACGCGCGGTGTCGACGCGCTGGTC 776
Qy 415 CGCGCAGCGCGCACCATC- --GTGGCAAGACGACCTCGAGGACATGGCGATGGTATC 471
Db 775 GAGGCGGCGCGGTGATCATCGCCGACACACGACCGCCGATTTCTCTGGTCCCTTC 716
Qy 472 GGTGAGCGACGCTTACGCTGCTGCTGTAACCCGACACACCCCGCCACGCGCGGT 531
Db 715 TGTCACTCGCGGCGATGCGGGGTGACGCGCAACCCGTAACCTTCGCGCTTCAAGCGCGC 656

```

```

Qy 532 GATCTTCCAGGGCTCGGGCTGCGCTCGCTCGCTCGCTCGCTCGCTCGCTCGCTCGCT 591
Db 655 GCTCTCTCGGGCGGCTCGGGGCGGCGCTCGGGGCGGCGCTCGGGGCGGCGCTCGGGC 596
Qy 592 GTGATGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 651
Db 595 TCGGACATCGGGCGGCTCGATTTCGGATCCGCGCTCGGCTCGGGGTGGTGGGTTCAG 536
Qy 652 GGGACCCAGCGGCTGCTGCGCTCTTACGGCTGACATACATGAGGAGGAGGAGGAGG 711
Db 535 CGCGCTTACGGCGGCTGCTGCGCTCGAC- --CGCGCGCTTCAACCTCGATCCCTAC 479
Qy 712 ATCGGGCGGCTCACCAGGGGGGCTCGAGCTCAACCGCGGGTCTCGAGGTCTTGGCC 771
Db 478 CGGGCGGCTATGGCGCGCACGGTTCGAGATTGCTGCTGATGTCAGAACGTCGCGCG 419
Qy 772 GCGGACTGCGGCGGAGCCCTAGTGGTGTGTAACCTTCGGGAGCGGAGAGTACGGTCC 831
Db 418 CCCCATCCGGAAGACATCGCTCGTGAACCCAGCTCGAACTCGGCTCGCGCTCGCGG 359
Qy 832 GCGCT 836
Db 358 GGGT 354

RESULT 13
LOCUS AZ934731
DEFINITION AZ934731.1 GI:13776791
ACCESSION AZ934731
VERSION AZ934731.1
KEYWORDS GSS.
SOURCE Bradyrhizobium japonicum
ORGANISM Bradyrhizobium japonicum
Bacteria; Proteobacteria; Alphaproteobacteria; Rhizobiales;
Bradyrhizobiaceae; Bradyrhizobium.
REFERENCE 1 (bases 1 to 521)
AUTHORS Tomkins,J.P., Wood,T.C., Stacey,M.G., Loh,J.T., Judd,A.,
Goicoechea,J.L., Stacey,G., Sadowsky,M.J. and Wing,R.A.
A marker-dense, sequence-ready map of the Bradyrhizobium japonicum
genome
JOURNAL Genome Res. 11 (8), 1434-1440 (2001)
MEDLINE 21376150
PUBMED 11483585
COMMENT Contact: Wing RA
Clemson University Genomics Institute
Clemson University
100 Jordan Hall, Clemson, SC 29634, USA
Tel: 864 656 7288
Fax: 864 656 4293
Email: rwing@clemson.edu
Class: BAC ends
High quality sequence stop: 482.
Location/Qualifiers
1. .521
/organism="Bradyrhizobium japonicum"
/mol_type="genomic DNA"
/strain="USDA110"
/db_xref="taxon:375"
/lab_hosts="E. coli"
/clone_lib="B. japonicum BAC library"
/note="Vector: pindigo536; Site 1: HindIII"

ORIGIN
Query Match 6.7%; Score 106.2; DB 28; Length 521;
Best Local Similarity 53.2%; Pred. No. 6.1e-06;
Matches 272; Conservative 0; Mismatches 233; Indels 6; Gaps 2;

```

```

Qy 270 GGGCGCCAGGAGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCG 329
Db 11 GGATGCGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 70

```

330	QY	CGCCGGTATGCCCAACACAGGAACGGGTCCCGGATGCTCCGCACTGTGATCGCCACCGAGGA	389
71	Db		
71	Db	GCTGGCTTTCGACACACG--GCGGCTCCCGGCTTCTCTACACACCGACCATGA	127
330	QY	TGCCGTGTGTGAGAGGGGTGCTCGCGGAGAGCGCACCATCGTCGGCAAGACAAACT	449
128	Db		
128	Db	CTCAGACGCGTGTGAGCGCTCGCGCGCGCGGCCATCATCATCGGCAAGACCAATCT	187
450	QY	CGAGGACATGGGATGG--TATCGGTGAAGGACAGGTTACGGTCTCGCGCTGAACCC	506
188	Db		
188	Db	CGACAGTTCCGAACCGGCTCTGTGCGGCTACGCTGCGCTTACGGCATTCCTCAAAATTC	247
507	QY	GAACAACCCGCCACAGCGTGTGATTTACAGCGCTCCGCGCTGCGCTGCTGCTGC	566
248	Db		
248	Db	GATTCGAGGATCTGATTTCCGGTGGCTCCAGTCTCGGCTCGGACCGCTCTCGGCG	307
567	QY	CGGCATGCTCGACTTCGCCCTGGCGCTCGATGAGGACGAGCAGCATCCGGATCCCGCGCG	626
308	Db		
308	Db	CGGCTGTGTGCGGTGACGCTCGGCAACGACAACCGCGGAGCGGCGCGGTGCCGCGCAT	367
627	QY	ATGTGTGGGACTGTGTGGCATGAAGGCGCACCAACGCGCTGTGTGCGCTTTACGGGCTGAC	686
368	Db		
368	Db	GCTCAACAACATCGTCGGGCTGAAGCAAGCCTCGGCATGATCTCACACGCGGACTCGT	427
687	QY	ATACATGGACACACCTTGGACACATCTGGGSCCATCACAGGGGGTGGAGCTCAACGC	746
428	Db		
428	Db	GCCTGCCCTGCCGACCTCGACTCGATCTGGGATTCGCCCTTACCGGTGGACGATGCTGC	487
747	QY	CCGGTCTCTGAGGTGTGGCGGGGCGAC	777
488	Db		
488	Db	ACTCGCACTCTCCGTGATGGCGGACCTGAC	518

```

RESULT 14
  AZ934200/c
  LOCUS
  DEFINITION
  ACCESSION
  VERSION
  KEYWORDS
  SOURCE
  ORGANISM
  REFERENCE
  AUTHORS
  TITLE
  JOURNAL
  MEDLINE
  PUBMED
  COMMENT
  FEATURES
    source
      672 bp DNA linear GSS 24-APR-2001
      BJ_Ba0002A21r B. japonicum BAC library Bradyrhizobium japonicum
      genomic, genomic survey sequence.
      AZ934200
      AZ934200.1 GI:13776260
      GSS.
      Bradyrhizobium japonicum
      Bradyrhizobium japonicum
      Bacteria; Proteobacteria; Alphaproteobacteria; Rhizobiales;
      Bradyrhizobiaceae; Bradyrhizobium.
      1 (bases 1 to 672)
      Tomkins,J.P., Wood,T.C., Stacey,M.G., Loh,J.T., Judd,A.,
      Golcochea,J.L., Stacey,G., Sadowsky,M.J. and Wing,R.A.
      A market-dense, sequence-ready map of the Bradyrhizobium japonicum
      genome
      Genome Res. 11 (8), 1434-1440 (2001)
      21376150
      11483585
      Contact: Wing RA
      Clemson University Genomics Institute
      Clemson University
      100 Jordan Hall, Clemson, SC 29634, USA
      Tel: 864 656 7288
      Fax: 864 656 4293
      Email: rwing@clemson.edu
      Class: BAC ends
      High quality sequence start: 24
      High quality sequence stop: 627.
      Location/Qualifiers
        1..672
          /organism="Bradyrhizobium japonicum"
          /mol_type="genomic DNA"
          /strain="USDAL10"
          /db_xref="taxon:375"
          /lab_host="E. coli"
          /note="Vector: pIndicos36; Site 1: HindIII"

```

FEATURES

```

/organism="Bradyrhizobium japonicum"
/mol_type="genomic DNA"
/dbname="USDA110"
/db_xref="taxon:375"
/lab_host="E. coli"
/clone_lib="B. japonicum BAC library"
/notes="Vector: pBrd19Q536; Site 1: HindIII"

```

ORIGIN

Query Match	6.7%;	Score 106;	DB 28;	Length 672;
Best Local Similarity	57.3%;	Pred. No. 7e-06;		
Matches 212;	Conservative 0;	Mismatches 155;	Indels 3;	Gaps 1;
257	TCTGCGGGTGGAGGGCGGCACGAGGGGCCCATGTAGCGACCTGTAGCGCGCGCATCAAGG	316		
425	TCTGCGATCGCGCGCCGCGCGCATCGGACCGCTGCATGGCGTGCCGATCGGGATCAAGG	366		
317	ACTGATCCGCATTCGCGCGTATGCCCCACACAGAACGGGTCCCGGATGCTCCCGACTGTGA	376		
365	ACCTCACCATATACCGCGGGCTGACCACCACCTACGGCTCGGCTTTGTTCCGCGATCAGC	306		
377	TGCGCACCGAGATGCGGTGGTGGAGCGGTGCTCGCGGACGCGCCACCATCGTCG	436		
305	TCCGGCCGAGACGAGCTCTGTGTCGCGGCTGCGGCGCGCGCGGCGATCATCTCG	246		
437	GCAAGACGAACCTCGAGGACATGGCGATGGGT---ATCGGTGAAGCAGCGGTCTACGGTC	493		
245	GAAGACCAACACGCCGGAATTCGGTTCGGTCTGCACCAACCGGTTGCGCGGGC	186		
494	CTGCGCTGAACCGGAACAACACCGCGCCACGGCACGGGTGGATCTTCAGCGGGTCCGGCG	553		
185	CGACGGCGCAATCCCTTCGATCCCGCGCTGACCTCCGGCGGATCGTCCGGCGGTCTCCCG	126		
554	CTGCGCTCGCTGCGCGCATGTGCGACTTCGCCCTCGGCGGTTCGATGAGGACAGGACGATCC	613		
125	TTGCGGTACGCCCGCGCATGAACCGCTGGCGACCGCACCCGATTTCCGGCGGCTCGGTGC	66		
614	GGATCCCGGC	623		
65	GCACGCCCGC	56		

RESULT 15

LOCUS	DEFINITION
EG300283	

BG3000285 517 bp mRNA linear EST 22-OCT-2001
 HVSM8E00011P07f Hordeum vulgare seedling shoot EST library
 HVCDNA0002 (Dehydration stress) Hordeum vulgare subsp. vulgare cDNA
 clone HVSM8E00011P07f, mRNA sequence.

ACCESSION

VERSION
KEYWORDS

SOURCE ORGANISM

1000

REFERENCE AUTHORS

TITLE

TO: DIRECTOR

COMMENT

FEATURES
SOURC

```
/db_xref=taxon:112509"
/clone="HVSMEB0011P07"
/tissue_type="Seedling shoot"
/lab_host="TJCI21"
HVCNDA0002 (Dehydration stress)"
/notes="Vector: lambdaZAP; Site 1: EcoRI; Site 2: XhoI;
Seeds were surface sterilized then germinated under axenic
conditions in the dark at room temperature on filter paper
with water, nystatin and cefotaxime in covered
crystallization dishes. Five-day old seedlings were
incubated at 90% RH for 24 hr. Shoots were then harvested,
total RNA was prepared, poly(A) RNA was purified, one
primary amplified cDNA library was made, 600000 pfu were
in vivo excised to give pBluescript SK(-) cDNA phagemids.
These steps were performed in the TJ Close laboratory at
the University of California, Riverside (Choi, Close,
Fenton). Phagemids were plated and picked at the Clemson
University Genomics Institute (CUGI) (Begum, Palmer,
Frisch, Atkins and Wing). Plasmid DNA preparations, DNA
sequencing and sequence analysis were performed at CUGI
(Wing, Yu, Frisch, Henry, Simmons, Oates, Rambo, Main).
The sequence has been trimmed to remove vector sequence
and contains a minimum of 100 bases of phred value 20 or
above. For more details on library preparation and
sequence analysis see
http://www.genome.clemson.edu/projects/barley. To order
this clone see http://www.genome.clemson.edu/orders Also
see Close TJ, Wing R, Kleinhofs A, Wise R (2001)
Genetically and physically anchored EST resources for
barley genomics. Barley Genetics Newsletter 31:29-30.
(http://wheat.pw.usda.gov/ggpages/bgn/31/cover.html)"
```

ORIGIN

```
Query Match          6.6%; Score 104.6; DB 12; Length 517;
Best Local Similarity 58.1%; Pred. No. 1e-05;
Matches 226; Conservative 0; Mismatches 154; Indels 9; Gaps 2;

QY 262 CGGTTGGAGGCGGCACGAGGGGCGCCTGACGCGGCTGCGCGGATCAAGGACTGC 321
Db 74 CAGGAGCAGTACCCCTGTCACGCGCTCACCTTCGCCATCAGGACATTTGACGTGCC 133

QY 322 ATGCCCATCGCGGTATGCCACCAACGCGGTCCCGGATGCTCCCGACTGTGATCGCC 381
Db 134 GGCGCGTCAACAGGTTCCGACCCCGGACTGGCGCCGAGCAGCGCGCGCCACC 193

QY 382 ACGAGGATCGGTGTGTGTGAGCGGTGCTCGCGGAGGCGCCACCATGTCTGGCAAG 441
Db 194 ACGGCGCCCGCGTGTGCGCCG-----GCTCGCGCGCGCGCCACCGGGGTGGGCATG 247

QY 442 AGCAACTCGAGACATGGCGATGGGTATCGGTGAAGCAGCGT---CTACGGTCTGCG 498
Db 248 ACGGTGATGACAAGATGCGCTACAGCATCACGGCGAGAACGCGCACTACGGCACCCCC 307

QY 499 CTGAACCCGAAACAAACCCCGCCACGCGGATCTTCAGCGGTCTCGGCGCTGCC 558
Db 308 GCCAACCCCTGGCGCCCGCGCGGTCCCGCGGTTCTCCAGCGGTTCGCGCGTCCGCC 367

QY 559 GTGCTGCGCGGATGTGCTACCTTCGCGCTGGCGTGTGATGAGCGAGGAGCATCGGATC 618
Db 368 GTGCGCGCGAGCTCGCCACATTCGCGCTGGGACCAACACCGCGCGGACGCTAAGGTTG 427

QY 619 CCGGCGCGATGTGCGGACTGTCGGCAT 647
Db 428 CCCCCACATCTCGCGGATTTTGGCCT 456
```

Search completed: June 27, 2004, 09:59:18
Job time : 4534.69 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: June 27, 2004, 04:34:02 ; Search time 126.584 Seconds
(without alignments)
6918.029 Million cell updates/sec

Title: US-10-658-691-4
Perfect score: 1578
Sequence: 1 ggcgacccaatgcgcatt.....ggcaccgggtgcggtag 1578

Scoring table: IDENTITY_NUC

Gapop 10.0 , Gapext 1.0

Searched: 682709 seqs, 277475446 residues

Total number of hits satisfying chosen parameters: 1365418

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

Issued Patents NA:**

- 1: /cgn2_6/ptodata/2/ina/5A COMB.seq.*
- 2: /cgn2_6/ptodata/2/ina/5B COMB.seq.*
- 3: /cgn2_6/ptodata/2/ina/6A COMB.seq.*
- 4: /cgn2_6/ptodata/2/ina/6B COMB.seq.*
- 5: /cgn2_6/ptodata/2/ina/PTUS COMB.seq.*
- 6: /cgn2_6/ptodata/2/ina/backfiles1.seq.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match %	Length	DB ID	Description
1	224.8	14.2	1521	1	US-08-726-136-20
2	224.8	14.2	1521	3	US-08-103-434-20
3	224.8	14.2	1521	3	US-08-687-594-20
4	180.8	11.5	1879	1	US-08-539-866-1
5	172.8	11.0	4403765	3	US-09-103-840A-2
6	170.4	10.8	4411529	3	US-09-103-840A-1
7	168.8	10.7	1878	1	US-07-612-673-1
8	168.4	10.7	1879	1	US-07-796-361A-10
9	151	9.6	1467	4	US-09-252-991A-9752
10	151	9.6	1818	4	US-09-252-991A-10084
11	123.4	7.8	4403765	3	US-09-103-840A-2
12	123.4	7.8	4411529	3	US-09-103-840A-1
13	120	7.6	2456	1	US-08-233-146-8
14	120	7.6	2456	1	US-08-463-470-8
15	114.6	7.3	1335	4	US-08-252-991A-15844
16	114.6	7.3	1746	4	US-09-252-991A-15790
17	114.6	7.3	3036	4	US-09-252-991A-15848
18	110.4	7.0	23673	4	US-09-773-816-1
19	108.8	6.9	1843	1	US-07-948-023-1
20	104.2	6.6	1209	4	US-08-252-991A-9930
21	102.2	6.5	4471	4	US-08-252-991A-15815
22	100	6.3	1401	4	US-09-489-039A-2359
23	93.8	5.9	1932	4	US-09-252-991A-10046
24	91.8	5.8	1395	4	US-09-252-991A-12734
25	91.8	5.8	1434	4	US-08-252-991A-12585
26	90.8	5.8	1462	3	US-08-961-083-41
27	90.8	5.8	1462	4	US-08-536-784-41

Query Match 14.2%; Score 224.8; DB 1; Length 1521;

28 90.8 5.8 1476 4 US-09-252-991A-7653 Sequence 7653, Ap
29 90.8 5.8 1485 4 US-09-252-991A-7738 Sequence 7738, Ap
30 90.8 5.8 2997 4 US-09-252-991A-7831 Sequence 7831, Ap
31 90.8 5.8 19702 4 US-08-961-527-7 Sequence 7, Appli
32 89.4 5.6 1401 1 US-08-726-136-27 Sequence 27, Appli
33 88.4 5.6 1401 3 US-09-103-434-27 Sequence 27, Appli
34 88.4 5.6 1401 3 US-08-687-594-27 Sequence 27, Appli
35 87.4 5.5 1640 1 US-07-796-361A-12 Sequence 12, Appli
36 87.4 5.5 1816 1 US-08-539-866-3 Sequence 3, Appli
37 87.4 5.5 1817 1 US-07-612-673-3 Sequence 3, Appli
38 80.4 5.1 44377 2 US-08-804-227C-7 Sequence 7, Appli
39 80.4 5.1 44377 2 US-08-804-198-1 Sequence 1, Appli
40 78.4 5.0 1230025 4 US-09-198-452A-1 Sequence 1, Appli
41 76 4.9 876 4 US-09-134-000C-1186 Sequence 1186, Ap
42 68.6 4.3 1476 3 US-08-896-346-1 Sequence 1, Appli
43 67.2 4.3 1347 4 US-09-107-532A-308 Sequence 308, App
44 66.2 4.2 2045 3 US-08-743-168B-42 Sequence 42, Appli
45 66.2 4.2 2045 3 US-08-743-168B-44 Sequence 44, Appli

ALIGNMENTS

RESULT 1

US-08-726-136-20
; Sequence 20, Application US/08726136
; Patent No. 5811286

GENERAL INFORMATION:
; APPLICANT: ROBERT D. FALLON
; APPLICANT: MARK S. PAYNE

APPLICANT: MARK J. NELSON

TITLE OF INVENTION: NUCLEIC ACID FRAGMENTS ENCODING
; TITLE OF INVENTION: STEREOSPECIFIC NITRILE HYDRATASE AND AMIDASE ENZYMES AND

TITLE OF INVENTION: RECOMBINANT ORGANISMS EXPRESSING THOSE ENZYMES USEFUL FOR
; TITLE OF INVENTION: THE PRODUCTION OF CHIRAL AMIDES AND ACIDS

NUMBER OF SEQUENCES: 28

CORRESPONDENCE ADDRESS:

ADDRESSEE: E. I. DU PONT DE NEMOURS AND COMPANY
; STREET: 1007 MARKET STREET
; CITY: WILMINGTON

STATE: DELAWARE

COUNTRY: UNITED STATES OF AMERICA

ZIP: 19898

COMPUTER READABLE FORM:

MEDIUM TYPE: FLOPPY DISK

OPERATING SYSTEM: MICROSOFT WINDOWS 3.1

SOFTWARE: MICROSOFT WORD 2.0C

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/08/726,136

FILING DATE:

CLASSIFICATION: 530

PRIOR APPLICATION DATA:

APPLICATION NUMBER: 60/004914

FILING DATE: OCTOBER 6, 1995

ATTORNEY/AGENT INFORMATION:

NAME: FLOYD, LINDA A.

REGISTRATION NUMBER: 33,692

REFERENCE/DOCKET NUMBER: CR-9677

TELECOMMUNICATION INFORMATION:

TELEPHONE: 302-892-8112

TELEFAX: 302-773-0164

INFORMATION FOR SEQ ID NO: 20:

SEQUENCE CHARACTERISTICS:

LENGTH: 1521 base pairs

TYPE: nucleic acid

STRANDEDNESS: single

TOPOLOGY: linear

MOLECULE TYPE: DNA (genomic)

HYPOTHETICAL: NO

US-08-726-136-20

Best Local Similarity 48.8%; Pred. No. 7,8e-35;
Matches 731; Conservative 0; Mismatches 752; Indels 15; Gaps 4;

QY 37 CGGACAGCCGACAGCTCCAGAGGTACAGCGCCGCCACCACTTCGACCTCGACGAGGAA 96
DB 16 CTAACCTCGACAGGTTTACAGATCCGAACCCAGTTGACATCACTCACTGACGACGAA 75
QY 97 GTGGCGGCCAGCTGTTCCGGTCTGTCGGGAGATGGTACTGCTTCGACCTGATCGAC 156
DB 76 CAGGAGAGTCTACTGGAAGTATGCAACGAGTTTCGAGCTTACGAGCTTGGTGCAC 135
QY 157 GAATACCGCAACCCCGCAGCCGCGACCGCGTACACGACCGCGCATCGGCCGAA 216
DB 136 GAACTGGGTGATTTGGTTCGGGCAATACGCTACGACCGCAGTTCAAGGTATCGCCATCG 195
QY 217 CGGACCGGCGAGAGACCCGTTCAACGATTCATCCGTTTCGCGGGTGGAGGGGCC 276
DB 196 CCATCGGCCAGGAAACCCCTGAAACGCTGTTACTACGAAACAGAAAGTAAAGTGGTGC 255
QY 277 ACGGAGGGCCACTGAGCGACCTGACCGCCGCGATCAAGGACTGCGATCGCCATCGCCGT 336
DB 256 CGCGAAGGCTCTGCGCGGCAAAACCGTCCGCTCAAAGATAATATCTCCTCGCAGGC 315
QY 337 ATGCCACACAAAGGTTCCGGATGCTCCGACTGTGATCGCACCGAGATCCGCTG 396
DB 316 GTCCCATGATGAACGGCGCAGCGCGTGGAGGCTTCGTCGCGGGTTCGATGCCACG 375
QY 397 GTGTGGAGCGCTGCTCGCGCAGCGCCACCATCGTCGCGCAAGACCACTCGAGGAC 456
DB 376 GTGGTCAACCGCTTGTCTGATCGCGGGGCGACCATCTCGCAAGACCCACCTCGGAGCAC 435
QY 457 ATGGCGAT---GGGTATCGGTGAAGGAGCGTCTAGGTCTGCGTGAACCCGAAACAC 513
DB 436 TACTGCTTTTGAAGGAGGACGACACCTCCGATCCAGCCCGCGTGCACACCCACATCGC 495
QY 514 CCGCGCCACGCGACCGGTGGATTTCCAGCGGCTCCGCGCTCGCTGCGTCCGCGCATG 573
DB 496 CACGGTTATGCTCTGGCGTTCTCATCAGGACGCGGCAATGTTGGTCCGCTGAG 555
QY 574 GTGACTTCGCTCGGCGTGCATGAGGAGGACGACATCCGATCCCGCGCATGTCG 633
DB 556 GTGGACATCGCTGGCGGCGATCAAGCGGCTCCATTCGATCCCGCTCGCGCTTCG 615
QY 634 GAGCTGGTGGATGAAGCGACCAACCGGCTGGTGGCTTACGGCTGACATACATG 693
DB 616 GGTACCTACGATGAGCCACCCACCGCTGGTGCCTACCGCGTATGCGCAT 675
QY 694 GACACACCTTGAACACATCGGCGCATCACAGGGGCTGAGCTCAACCGCGGTC 753
DB 676 GAAGCCACGATGATCATGTCGCGCCCATCACCGGTAAAGTGGCGGCAACGCGCTGATG 735
QY 754 CTCGAGGTGTGGCGGCGGCGACTGGCGCGACCTCATGTGGTGGTAACTTCGCGAG 813
DB 736 CTCGAGGCAATGGCGGTGCAGCGGACTGACCGCGCGCGCGGCGG-----CTCAG 789
QY 814 CCGGAGAACTAGCTTCGCGTTCGCGGAGGAGTATCGGTCTCAGATTCGCGTCTGTC 873
DB 790 GTGATGACTATTCAGTTTACCTGGAAGAGCGGTGAGCGGACTCAGAATCGGGTGTG 849
QY 874 GAGGAGTCACTGAGCGCAACGCTGCGACCGCGGCTGATCGCGCGTTCACCCAGGA 933
DB 850 CHAGAGGATTCGCTTCTAACAGGACCTTCGCTGGCGGCAAGTGGCGGACGCC 909
QY 934 CTGGCGGCGCTCGAGAGCGCGTGCACCATCGAGCGGCTTCGCTGCGGTGTGGACG 993
DB 910 ATCGCGGACTCGAGCGGTGGCGGTCTATGTCGAGCGGCTCTCCATTCGCGGACACA 969
QY 994 GCGGCTGCGCTATCAGAGCGCGTGTGGCTTTCAACCGCGCGCTATGCGGACTCC 1053
DB 970 CTGGCAGGTTGTGTGGCACCCCATTCGCTTGGAGGCTTGACCATGATGATGAT 1029
QY 1054 GCGGTTGGGTACTTTCACAGGCGCGCTGGAGCTCAGCACCGCGCTCAGACGCGG 1113

1030 GGCAACGGCGCAGGCTTTAACTGGAAAGGACTTTACGATGTGCGGCTGCTGGACAAACAA 1089
1114 GCGCAGAGTCGACCAACCAAGGATCTGGCGATCTCTCCGGGTGATGCTGGTATC 1173
1090 GCAGCTGCGCGACGACGACCAATATATCCGCTGCTCAAGCTCTGCTGCTGTC 1149
1174 GCGGACACCTGCGCGACGAATACCTCGGCATCTCACTAGCGAAGCGCAGAACCTGCGG 1233
1150 GGCAATAGGCTGTCGCGCTACAACGACGCTACTAGCCCAAGGCCAGAACCTTGCA 1209
1234 CTGGAGCTCGCAAGCAGATCGAGCGCTCTCCAGGACCGGGCTGCACTGTGACCCCG 1293
1210 CGCTTTCGCGCGGAGGATACGACAAAGCGCTGCAAACTATGACCTGCTGGTATCGG 1269
1294 ACCAGCCTACGTTGCCAACGAGCTGTTGAGCGGTGCGCAAGACCACTATGCTCATGATC 1353
1270 ACCAGCCTACGTTGCCAACGAGCTGTTGAGCGGTGCGCAAGACCACTATGCTCATGATC 1326
1354 CCAGGATGACGGGCAATCGGATCTCAACAGTGTGCGCGCTGCACTACCGGTACCCG 1413
1327 GTGGCTCGCGGCTTGAAGATGATCGGCAATACCGCGCCACAGGACATCACCGGATCGG 1386
1414 GCCTCAGCGTGCACCGGTCGCGGCGGAGAGGCGCTCGCGCTTGGCTTCAAGTGATA 1473
1387 GCATGTGATTC---CGTGGCTGCTGAGCGCTTCCGCTCGGGTGTGCTGCTGTC 1443
1474 GGCGGCACTTCGAGGAGTCAAGCTTACCGCACCGCGCGCTGATCAGGCGCGCG 1531
1444 GCAAAACACTACCGCGGAGGACGATTACCAAGCGCGCGCGCTTGAAGCGCTCGG 1501

RESULT 2
US-09-103-434-20
; Sequence 20, Application US/09103434
; Patent No. 6133421
; GENERAL INFORMATION:
; APPLICANT: ROBERT D. FALLON
; APPLICANT: MARK S. PAYNE
; APPLICANT: MARK J. NELSON
; TITLE OF INVENTION: NUCLEIC ACID FRAGMENTS ENCODING
; TITLE OF INVENTION: STEREOSPECIFIC NITRILE HYDROLASE AND AMIDASE ENZYMES AND
; TITLE OF INVENTION: RECOMBINANT ORGANISMS EXPRESSING THOSE ENZYMES USEFUL FOR
; TITLE OF INVENTION: THE PRODUCTION OF CHIRAL AMIDES AND ACIDS
; NUMBER OF SEQUENCES: 28
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: E. I. DU PONT DE NEMOURS AND COMPANY
; STREET: 1007 MARKET STREET
; CITY: WILMINGTON
; STATE: DELAWARE
; COUNTRY: UNITED STATES OF AMERICA
; ZIP: 19898
; COMPUTER READABLE FORM:
; MEDIUM TYPE: FLOPPY DISK
; COMPUTER: IBM PC COMPATIBLE
; OPERATING SYSTEM: MICROSOFT WINDOWS 3.1
; SOFTWARE: MICROSOFT WORD 2.0C
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/103,434
; FILING DATE:
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US/08/726,136
; FILING DATE:
; APPLICATION NUMBER: 60/004914
; FILING DATE: OCTOBER 6, 1995
; ATTORNEY/AGENT INFORMATION:
; NAME: FLOYD, LINDA A.
; REGISTRATION NUMBER: 33,692
; REFERENCE/DOCKET NUMBER: CR-9677
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 302-892-8112
; TELEFAX: 302-773-0164
; INFORMATION FOR SEQ ID NO: 20:

SEQUENCE CHARACTERISTICS:
LENGTH: 1521 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: DNA (genomic)
HYPOTHETICAL: NO
US-09-103-434-20

Query Match 14.2%; Score 224.8; DB 3; Length 1521;
Best Local Similarity 48.8%; Pred. No. 7.8e-35;
Matches 731; Conservative 0; Mismatches 752; Indels 15; Gaps 4;

QY 37 CCACAGCCGACAGCTCCAGAGGTACAGCGCCGACCACTTCGACCTCGACGAGAA 96
DB 16 CTTACCTCGACCCAGGTTTATGACATCCGAACCCAGTTTCGACATGCAACTGACGACGAA 75
QY 97 CTGGCCGCCAGCTCGTTCCGGTCGCGGAGATGTTGACTGCTTCGACCTGACCTGATCGAC 156
DB 76 CAGGACAGCTCTACTCGAATGATGCAACCGAGTTTCGACGCCCTACGACCTGCTGCAC 135
QY 157 GAACTACCCGAACCCGCCGAGCCGCGACGCCGTACACGGAACCGGACATCGCCCGGAA 216
DB 136 GAACTGGCTGATTCGTTCCGCGCAATACGCTACGACCGCAGTTTCAGGCTATCGCCATCG 195
QY 217 CCGACCGCGACAGACCGGTTCAACGCATTCATCCGGTTCTGCGGGTGGAGGCGCC 276
DB 196 CCATCGCCGCAAGAAACCTCTGAACGCTTGTACTACCGAACAGAGTGAATGTGTC 255
QY 277 ACGGAGGGGCACTGACGACCTGACCGCGGATCAAGGACTGCATCCCATCGCCGT 336
DB 256 CGCGAAGGCTGTGCGGGGCAAAACCGTCGCGCTCAAGATAATATCTCCCTGGCAGCG 315
QY 337 ATGCCACACGAAACCGGTCGCGAGTGTCCCGACTGTGATGCCACCGAGGATCCCGT 396
DB 316 GTCCCATGATGAACGCGCAGCGCGCTTGAAGGTTCTGTCGGGGTTGATGTCACG 375
QY 397 GTGGTGGAGCGGTGTGCGGGCAGCGCCACCATCGTGGCAAGACCACTTCGAGGAC 456
DB 376 GTGGTCACCGCTTGTCTGATGCGGGGCGACCATCTCGGCAAGCCACCTGCGAGCAC 435
QY 457 ATGGCAT---GGTATCGTGAAGGAGCGGTCTACGGTCTCGCTGTAACCGGAACAC 513
DB 436 TACTGCTTTTCAGGAGCACCACACCTCGCATCCAGCCCGGTGACACACCCACATCG 495
QY 514 CCGCCCAACGCACTGAGTCTTCAGCGGCTCGCGGCTCGCTGCTGCTGCGCGCATG 573
DB 496 CACGGTTATGCTTCTGCGGTTCTCTATCAGGACGCGGCTATGTTGGTCCGTGAG 555
QY 574 GTGACTTCGCTTGGCGTGTGATGAGGACGAGCATCCGATCCCGGCGCATGGTGC 633
DB 556 GTGACATCGCGTGGCGCGATCAAGCGGCTCCATTCGATCCGCTCGCGCTTCTGC 615
QY 634 GACTGTCGCGATGAAGGCGACCCACGGCTGCTGCTGCTTACGGCTGACATACATG 693
DB 616 GGTACCTACGATGAAGCCACCCAGCGCTGTGCTTACACCGCGTTCATGGCGATT 675
QY 694 GACCACCTTTGGACCATCGGCCCATCACCAGGGGGTTCGAGTCAACCGCGCGTC 753
DB 676 GAAGCCAGATCATGATGCGGCCCATCACCGTTAAGTGGCGGACACCGCTGATG 735
QY 754 CTCGAGTGTGGCGGGCGGCTGCGGACCTCAGTGGTGGTGAACCTTCCGGAG 813
DB 736 CTGAGGCAATGGCGGTGACAGCGGCTGACCCCGCGCGAGGGCGG-----CTTCAG 789
QY 814 CCGGAGAACTACGGCTCGCGCTCGGAGGAGTATCCGGTCTGAGATTCGGGGTGC 873
DB 790 GTCGATGACTATTGAGTTACTCTGAAAGAGGCGTGAAGGAGTCAAGATCGGGTGTG 849
QY 874 GAGAGTCACTGGAGCGGAACGGTCGACCGCGGACGCTGATCGCGGCTTCAACAGGGA 933
DB 850 CAAGAGGATTCGCGCTTGTCAACAGGACCTTCGCTGGCGGCAAAAGTGGCGAGCGC 909

QY 934 CTGGCGCGCTCGAGAGCGCGGTCGCAACCATCGAGCGGCTCTGGTGGCGTGTGAGCG 993
DB 910 ATGCCCGACTCGAGGCGTTGGCGCTCATGTGAGCGGCTCCATTCGAGCACAAC 969
QY 994 CGGCTTGGCTATCCAGAGCGCGGTGATGGCTTCAACGCGCGGCTATGGGGACTCC 1053
DB 970 CTGGCAGGGTGTGTGGCACCCCATCGGTGGAGGCTTGACCATGAGATGATCAT 1029
QY 1054 GCCGCTGGGCTACTTTCCAAAGGGGCGGTGAGCGTCAAGCACCGCGCTCAAGAGCGG 1113
DB 1030 GGCAAGCGCGCAGGCTTTAACTGGAAGGACTTTACGATGTCGCGCTCTGGAACAACAA 1089
QY 1114 GCCCAGAGTCCGACCAACCCACAGGATCTCGCATCTCTGCGGCTGATGCTGCTGATC 1173
DB 1090 GCCAGCTGGCGGACGACGAGCAATATTCGCGCTCAAGCTCTGCTGCTGCTGCTGCT 1149
QY 1174 GCGGAGCACCTGCGCGACGAATACCTCGCATCCACTACGCGAAGGCGGAGAACCTCGG 1233
DB 1150 GGCCAAATACGCGCTGTGCGCTACACGAGCGCTACTACGCAAGGCCAGAACCTTGCA 1209
QY 1234 CTGGAGCTGGCAAGAGATCGAGCGGCTCTCCAGGACCGGGCTGACCTGACCGCG 1293
DB 1210 CGTTTCCCGGCGAGGATACGAAAGCGCTGCAAAACCTATGACCTGCTGGTATCGCG 1269
QY 1294 ACCAGCGCTACCGTTGCAACGAGCTGTGAGCGGTCGCGAAGACACCATGTCCATGATC 1353
DB 1270 ACCAGCGCTACCGGCGCAACCCACCCCGCAG---CGAATGCTCGATCAGGAGTAC 1326
QY 1354 CCAGGATGAGCGGCAATGATCTCTCAACAGCTGCGCGCTGAGCTTACCGGTCACCGG 1413
DB 1327 GTGGCTCGCGGCTTGGAAATGATCGCAATACCGCGCACAGGACATCACCGGCGATCCG 1386
QY 1414 CGCTGACGCTGCGCACGCGGTGCGGCGAGAGGCGCTGCGGCTTGGCTTCAAAGTGATA 1473
DB 1387 GCAATGCTGATC---CGTGGGCTGCTGAGCGGCTTGGCGCTGCTGCTGCTGCTGCT 1443
QY 1474 GCGCGCATTCGAGGAGTGAAGCTTACGCAACCGCGCGGCTGATGAGGCGCGCG 1531
DB 1444 GCAAAACACTACGCGGCGGCGAGTATACCAAGCGCGGCGGCTTGAAGCGCTCGG 1501

RESULT 3

US-09-687-594-20
Sequence 20, Application US/09687594
Patent No. 6251650

GENERAL INFORMATION:

APPLICANT: ROBERT D. FALLON

APPLICANT: MARK S. PAYNE

APPLICANT: MARK J. NELSON

TITLE OF INVENTION: NUCLEIC ACID FRAGMENTS ENCODING

TITLE OF INVENTION: STEREOSPECIFIC NITRILE HYDRATASE AND AMIDASE ENZYMES AND

TITLE OF INVENTION: RECOMBINANT ORGANISMS EXPRESSING THOSE ENZYMES USEFUL FOR

TITLE OF INVENTION: THE PRODUCTION OF CHIRAL AMIDES AND ACIDS

NUMBER OF SEQUENCES: 28

CORRESPONDENCE ADDRESS:

ADDRESSEE: E. I. DU PONT DE NEMOURS AND COMPANY

STREET: 1007 MARKET STREET

CITY: WILMINGTON

STATE: DELAWARE

COUNTRY: UNITED STATES OF AMERICA

ZIP: 19898

COMPUTER READABLE FORM:

MEDIUM TYPE: FLOPPY DISK

OPERATING SYSTEM: MICROSOFT WINDOWS 3.1

SOFTWARE: MICROSOFT WORD 2.0C

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/09/687,594

FILING DATE:

CLASSIFICATION:

PRIOR APPLICATION DATA:

APPLICATION NUMBER: 08/726,136

FILING DATE:

ATTORNEY/AGENT INFORMATION:
NAME: FLOYD, LINDA A.
REGISTRATION NUMBER: 33,692
REFERENCE/DOCKET NUMBER: CR-9677
TELECOMMUNICATION INFORMATION:
TELEPHONE: 302-892-8112
TELEFAX: 302-773-0164
INFORMATION FOR SEQ ID NO: 20:
SEQUENCE CHARACTERISTICS:
LENGTH: 1521 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: DNA (genomic)
HYPOTHETICAL: No
US-09-687-594-20

Query Match 14.2%; Score 224.8; DB 3; Length 1521;
Best Local Similarity 48.8%; Pred. No. 7.8e-35;
Matches 731; Conservative 0; Mismatches 752; Indels 15; Gaps 4;

QY 37 CCGACAGCCGAAACAGCTCCAGGAGTACAGCGCCGCCACCACTTCGACCTCGACGAGAA 96
DB 16 CTAACCTCGACAGGTTTACAGATCCGAACCCAGTTGCACATGCAACTGACGACGAA 75
QY 97 CTGGCGCCCGACAGTCTGTCGGTCTGTCGGGAGATGGTACTGCTTCGACTGATCGAC 156
DB 76 CAGGAGAGCTCTTACTCGAACTGATGCAACCGAGTTTCGACGCTACGACCTGCTGAC 135
QY 157 GAACTACCCAAACCGCGCAGCGCGGCGGACGCGCTACAGGACGCGGACATCGGCGCGAA 216
DB 136 GAACTGGCTGATTCTGTTCCGCCAATACGCTACGACCGCAGTTTCAGGCTATCGCCATCGG 195
QY 217 CGACCGGCGCAGACAGCCGTTTAAAGCAATTCATCCGGTCTGTCGGGTTGAGGCGCC 276
DB 196 CCATCGGCCAAGGAAACCTCTGAACCGCTGCTACTACCGAAAGAAAGTGAATGGTCC 255
QY 277 ACGGAGGGCCACTGAGCACCTGACCGCGCGGATCAAGGACTGATCGCCATCGCGGT 336
DB 256 CGCGAAGGCTGCTGGCGGCAAAACCGTCCGCTCAAGATAATATCTCCCTGCGAGC 315
QY 337 ATGCCACACGAAACGGTTCGGGATGCTCCGATGTGTCGATCGCACCGAGATGCCGTG 396
DB 316 GTCCCATGATGAACGGCGCAGCGCGCTTGGAAAGGCTTCGTCGCCGGGTTCGATGCCACG 375
QY 397 GTGGTGGAGCGCTGCTCGCGCAGCGCGCCACATCGTCGCGCAAGCAACCTCGAGGAC 456
DB 376 GTGGTACCCGCTTCTGATCGGGGGGACCATCTCGGCAAGCCACTCTGCGAGCAC 435
QY 457 ATGGCGAT---GGGTATCGGTGAAGCGACGCTCTACGGTCTCTGCGCTGAACCCCGAACAC 513
DB 436 TACTGCTTTTACGAGGAGCGACACACCTCCGATCCAGCCCGGTGCACAAACCCACATCGC 495
QY 514 CCGCGCCACGACGCGGTGATCTTCCAGCGGCTCGGGGCTGCGCTGCTGCGCGCATG 573
DB 496 CACGGTTATGCTCTGCGGGTTCCTCATCAGGACGCGGGCATTTGGTTCGCTCGGTGAG 555
QY 574 GTGACTTCGCTCGCGCTGATGAGGAGGAGCAGCATCCGGATCCCGCGCGCATGTCG 633
DB 556 GTGGACATCGCGTGGCGCGCATCAAGCGGCTCCATTCGATCCCGTCCGCTTCTGC 615
QY 634 GGACTGGTGGGATGAGCGGACCCAGCGCTTGGTCCGCTTACGGCTGACATACATG 693
DB 616 GGTACTACGGATGAAGCCACCCACCGGCTGGTGCCTACACCGGGGTCTATGGCGATT 675
QY 694 GACCACACCTTGGACCAATCGGGGCCCATCACAGGGGGTTCGAGTCAACCGCGGGTTC 753
DB 676 GAAGCCACCATCGATCATCTCGGCCCATCATCGGTAACTGCGGCAACAGCGGTGATG 735
QY 754 CTCGAGGTGTTCCCGGGGCCGACTGGCGGACCTCTAGTGGGTGCGTAACCTTCCGAG 813
DB 736 CTCGAGGCAATGGCCGGTGCAGACGACTCGACCCCGCGCGCGCGC-----CCTCAG 789

814 CCGAGAACTACGGCTCCGGCTCGCGGAGGAGTATCCGGTCTGAGATTTCGGGTGTC 873
790 GTGATGACTATTTCAGTTACTTGAAGAGCGTGAAGCGACTCAGATCGGGTGTG 849
874 GAGAGTCACTGAGCGAAACGGTGGAGCGCGGAGTATGATCCCGCTTAAACAGGGA 933
850 CAAGAGGATTCGGCTTGTACTAACAGGACCTTCGGCTGGCGGACAAAGTGGCGACGC 909
934 CTGGCGCGCTCGAGAGCGCGGCGGACCATCGAGCGGGTCTCGGTGCGCTTGTGAGC 993
910 ATGCGCGACTCGAGGGTTGGCGCTCATGTGAGCGCGTCTCCATTCGAGGACAAAC 969
994 GCGGCTGCGCTATCCAGAGCGCGCTGATGGCTTTCAACGCGCGCTATGCGGATCC 1053
970 CTGGCAGGCTTGTGTGGCACCCCATCGGTTGGAAGGCTTGACATGATGATGAT 1029
1054 GCGGTGTGGCTACTTCCACAGGGGCGGTGGAGTCAAGCATCGACCGCTCAGACGGCG 1113
1030 GGCAACGGCGAGCTTTAACTGGAAGGACTTTACGATGTGCGCTGCTGCAAGCTCTG 1089
1114 GCCAGAGTGCACACCCCAACAGGATCTGGCGATCTCTGCCGGGTGATGTGCTGATC 1173
1090 GCCAGCTGGCGGACGAGCGACCAATATATCGCGCTGCTCAAGCTCTGCAATGTT 1149
1174 GCGAGCACTGCGCGACGATACCTCGGCATCCATAGCGAGGCGGAGACCTGCGG 1233
1150 GGCAATACGCGCTGTGCGGCTACAACGAGCGCTACTACGCCAAGGCCGAGACCT 1209
1234 CTGGAGTTCGGCAAGCAGATCGAGCGCGCTCTCCAGGACGCGGCTGCATCTGTGAC 1293
1210 CGTTTTCGGCGGAGATACGACAAAGCGCTGCAACCTATGACCTGCTGGTGGTGG 1269
1294 ACACGCTACGTTGCGAAGAGCTGTGAGCGGTGCGGAGAGACCAATGTCATGATC 1353
1270 ACCAGCCCATACGCGCCCAACCCCGCCAG---CGAACTGCTCGATCAGCGAGTAC 1326
1354 CCAGGATGACGGGCAATCGATCTCAACAGCTGCGCTGAGCTCACCGGTCAACCG 1413
1327 GTGCTCGCGGTGGAAATGATCGCAATACCGCGCACAGACATCACCGGCTCGG 1386
1414 GCGCTGACGTTGCCACGGGTGCGGCGGAGAGAGGCTGCGCTGCTGCGCTTCAAGTATA 1473
1387 GCCATGTGATTC---CGTGTGGCTCTGCTGAGCGGCTGCGCTGCGGTGATGCTG 1443
1474 GCGCGCACCTTCGAGGAGTCAAGCTCTACCGCACCGCGCGCTGATCAGAGCGCGCG 1531
1444 GCAAAACACTACCGCGAGGCGCAGTTACCAAGCGCGCGCGCTTGAAGCTCGG 1501

RESULT 4
US-08-539-666-1
Sequence 1, Application US/08539666
Patent No. 5766918
GENERAL INFORMATION:
APPLICANT: Petre, Dominique
APPLICANT: Cerbelaud, Edith
APPLICANT: Mayaux, Jean-Francois
APPLICANT: Yeh, Patrice
TITLE OF INVENTION: No. 5766918el Polypeptides, The DNA Sequences
TITLE OF INVENTION: Allowing their Expression, Method of Preparation, and
TITLE OF INVENTION: Utilization
NUMBER OF SEQUENCES: 17
CORRESPONDENCE ADDRESS:
ADDRESSEE: Finnegan, Henderson, Farabow, Garrett &
ADDRESSEE: Dunner
STREET: 1300 I Street, N.W.
CITY: Washington
STATE: DC
COUNTRY: USA
ZIP: 20005-3315
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible


```
; FILE REFERENCE: 24366-20007.00
; CURRENT APPLICATION NUMBER: US/09/103,840A
; CURRENT FILING DATE: 1998-06-24
; NUMBER OF SEQ ID NOS: 2
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 2
; LENGTH: 4403765
; TYPE: DNA
; ORGANISM: Mycobacterium tuberculosis
; FEATURE:
; OTHER INFORMATION: CDC 1551
; OTHER INFORMATION: "n" bases at various positions throughout the sequence
; OTHER INFORMATION: represent a, t, c or g
US-09-103-840A-2

Query Match      11.0%; Score 172.8; DB 3; Length 4403765;
Best Local Similarity 48.9%; Pred. No. 1.5e-24;
Matches 623; Conservative 2; Mismatches 619; Indels 30; Gaps 5;

QY 275 CCACGAGGGGCACTGACGACCTGACCGCGCATCAAGGACTGTCATCCCATCGCGC 334
Db 3365561 CTTTCCCGTCGCGCTGGCCGGGTGCGCTGGCGCTCAAGGACGTCTTACACACGCG 3365502

QY 335 GTATGCCACACGAAAGGTCCTCCGATGCTCCGATGTTGATCCGACGAGATGCG 394
Db 3365501 ACATGCCACACCTCGGGTCAAAATCTCTGGAGGATGCGATCTCCCTACGACGCA 3365442

QY 395 TGGTGTGAGCGGTGCTCGCGGAGGCGGCGGACCATCTGCGGCAAGACGAACTCGAGG 454
Db 3365441 CGCTGACCGCGGTGTCGCGCGGCGGGATCCGATCTCTGGGCAAGACCAATGACG 3365382

QY 455 ACATGCGATGGG---TATCGGTGAAGGAGCGTCTACGGTCTCTGCTGAACCGGACA 511
Db 3365381 AGTTGCGATGGGCTCGTCAAGGAGAACTCCGCTTACGGTCCCAACCGCAACCGTGA 3365322

QY 512 ACCCGCCACGCGCGGTGATCTTCCAGGGTCCGGCGCTCGCTCGTCCGCGCA 571
Db 3365321 ATCTGACCGGTATCCGCGGTTTCGGTGGCGGAGCGCGCGCGCTGCGCGGTTCC 3365262

QY 572 TGGTCACTTCGCTGGCGCTGATGAGGAGGAGGATCCGATCCGATCCGCGCGCATGGT 631
Db 3365261 AGGCGCGCTGCGCATGATGAGGAGGAGGATCCGATCCGATCCGCGCGCATGGT 3365202

QY 632 GGGGACTGTCGATGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 691
Db 3365201 CCGGACCGTGGGTCGATGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 3365142

QY 692 TGGACACACCTTGGACACATCGGCGCCCATCACAGGAGGAGGAGGAGGAGGAGGAGG 751
Db 3365141 GCGGTCCTCGCTGATCAGGCGGCGCGCTGTCGCGGACCGCTTTCGACACCGCGTGT 3365082

QY 752 TCTCGAGTGTGGCGGCGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCG 811
Db 3365081 TGCATCAGTGTGTCGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 3365022

QY 812 AGCGGAGAACTACGCTCGCGCTCGCGGAGGAGGATTCGCGTCTGAGATTCGCGGTCG 871
Db 3365021 CCGAGTGTGGCGCGCTAGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCG 3364962

QY 872 TCGAGAGTCTATGAGCGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 925
Db 3364961 GCGTGTTCGACAGTGTGACGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 3364902

QY 926 ACCAGGAGTGTGGCGGCGCTGAGAGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 985
Db 3364901 AGGTCGCTGAGAGGAGTATACCGGCGTGGCGGCTGAGGTGAGGAGGAGGAGGAGGAG 3364842

QY 986 TGTGACGCGCGCTGGCTATTCAGAGCGGCGGCTGATGCTTCAACG---CGCGCGCTA 1042
Db 3364841 ACTTCAGCATGCGCTGGCGGCTTATTCATGATTCGCTCGGAGGTGTCGAGCAATC 3364782

QY 1043 TGGCGACTCCGCGGTGGGCTACTTCCAAAGGGCGGTGAGCATGACGACCGCGCG 1102
Db 3364781 TGGCGCGCTTTCGACGCGGATGCGCTACGGGTCGGGTCGGGACGACGACCGCGCGCG 3364722

QY 1103 TCACGACGCGCGCGCGGATGCGCACCAACCAAGGATCTGGCGATCTCTGTCGCGCTGA 1162
Db 3364721 CCAGGAGGTGATGCGGATGACCCGGGCGCGCGGTTTCGGGCGCGAGGTCAAGCGCGCA 3364662

QY 1163 TGGTGTGATCGCGGAGACCTGCGGAGCAATACCTCGGCATCCACTACGCGAGGCGC 1222
Db 3364661 TCATGATCGGACCTACGGTGTGTCGGCGGCTACTACGACGCTATTACAAACGAGCGC 3364602

QY 1223 AGAACCTTGGGCTGAGGCTCGGCAACAGATGAGCGCGCTCTCTCCAGGACCGGGTTCAC 1282
Db 3364601 AGAAGTGGCAGCTGATCGCGCGGACCTCGACGCGGCGATCGGTCCGTCGAGTGC 3364542

QY 1283 TGGTGAACCGGACGACGCTTACGTTGCGCAACGAGCTGTGAGCGGTGCGGAGACCA 1342
Db 3364541 TGGTGTGCGCCACGACCGCGCGCTTCCGCGTGGGTGAGAGGTGAGGATCCCG 3364482

QY 1343 TGTCCATGATCCACGATGAGCGGCAATGCGATCTCAACAGCTGCCCGCTGGACCTCA 1402
Db 3364481 TGGCGATGATCTGTTGCA-----CCTGTGACGCTGCGCTGMACTTG 3364437

QY 1403 CCGGTCAACCGCGCGCTGACGTTGCCCAACGCGTG---CGGCGGAGAGGCGCTTCCCGCTTG 1459
Db 3364436 CCGGCACTGCGGCTGCTGTGCGCTCGGGCTCTCCCGGACGACGCGGTTCGCGGTTG 3364377

QY 1460 GCCTCAAGTGTAGGCGCGCACTTCGAGAGTGCAGCGCTCTACCGCAACCGGCGCGTGA 1519
Db 3364376 GCCTACAGATCATGGCGCGGCAATGCGCGACGACCGGCTCTACCGGTGGGCGCGCTT 3364317

QY 1520 TCAGGCGCGCGCGC 1533
Db 3364316 ATCAGGCGCGCGCGC 3364303

RESULT 6
US-09-103-840A-1/c
; Sequence 1, Application US/09103840A
; Patent No. 6294328
; GENERAL INFORMATION:
; APPLICANT: FLEISCHMAN, Robert D.
; APPLICANT: WHITE, Owen R.
; APPLICANT: FRASER, Claire M.
; APPLICANT: VENTER, John C.
; TITLE OF INVENTION: DNA SEQUENCES FOR STRAIN ANALYSIS IN MYCOBACTERIUM
; TITLE OF INVENTION: TUBERCULOSIS
; FILE REFERENCE: 24366-20007.00
; CURRENT APPLICATION NUMBER: US/09/103,840A
; CURRENT FILING DATE: 1998-06-24
; NUMBER OF SEQ ID NOS: 2
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 1
; LENGTH: 4411529
; TYPE: DNA
; ORGANISM: Mycobacterium tuberculosis
; OTHER INFORMATION: H37Rv
US-09-103-840A-1

Query Match      10.8%; Score 170.4; DB 3; Length 4411529;
Best Local Similarity 48.9%; Pred. No. 4.4e-24;
Matches 623; Conservative 0; Mismatches 621; Indels 30; Gaps 5;

QY 275 CCACGAGGGGCACTGAGGACCTGACCGCGCATCAAGGACTGTCATCCCATCGCGC 334
Db 3371229 CTTTCCCGTCGCGCTGGCGGGTTCGCGTGGCGTCAAGGAGCTTTCACCAACGCG 3371170

QY 335 GTATGCCACACGAAAGGTCCTCCGATGCTCCCGATGCTGATGTCGCAACGAGGATGCG 394
Db 3371169 ACATGCCACACCTCGGGTCAAAATCTCTGGAGGATGCGGATCTCCCTACGAGCGCA 3371110

QY 395 TGGTGTGAGCGGTGCTCGCGGAGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 454
Db 3371109 CGCTGACCGCGGTTGCGGCGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 3371050
```

QY 455 ACATGCGATGGG---TATCGGTGAAGGAGAGGTCTACGGTCTCGGTGAACCCGAAACA 511
Db 3371049 AGTTCGCGATGGCTCGTACAGGAACTCTCGCTTACGGTCCACCCGCAACCCGTGA 3370990

QY 512 ACCCGCCCAACGCGAGGTGATCTTCAGCGGCTCCGGCTCGCTCGCTCGCGCA 571
Db 3370989 ATCTCAGCGGTACCCGCGGTTCGCGTGGCGGACGCGCGGCTGCGCGGTTC 3370930

QY 572 TGTGACATTCGCGCTGGCGTTCGATGAGGACAGGACATCCGGATCCCGGCGCGATGT 631
Db 3370929 AGCGCGCTGCGCATCGATCGACACCGGCGGTTCGATCCGCCAGCGCGCGCTGA 3370870

QY 632 GCGGACTGTCGCGATGAAGCGACCCAGCGGCTGCTGCGCTTACGGCTTCGACATCA 691
Db 3370869 CCGCGACCGTGGCGGTCAAAACCCACTACGCGCGGTTCGCGCTATGCGCTGGTGGCT 3370810

QY 692 TGACACACACTTGGACCAATCGCGCCCATCACAGGGGGTTCGAGCTCAACGCCCGCG 751
Db 3370809 GCGCGTCTCTGCTGATCAGGCGCGCGCTGTGCGCGCACCGCTCTTGGACACCGGTGT 3370750

QY 752 TCTCAGAGTGTGGCGCGGCGCACTGCGCGACCTCAGTGGGTGCGTAACTTCGCG 811
Db 3370749 TGCATCAGGTGATCGCGCGCACAGACCGCGGACTCCACGTGCTGCGACGCGAGTGC 3370690

QY 812 AGCGGAGAACTACGCTCGCGCTCGCGAGGAGTATCCGCTCTGAGATTCGCGGTGC 871
Db 3370689 CCGAGTGTGGCGCGCTAGGCGCGCGCTGCGGTGCGGATCTGCGTGGCTGCGGTGC 3370630

QY 872 TCAGGAGTCACTGGAGCGCAACGCTGCG-----ACGCGGACGTGATCGCGCGTTCA 925
Db 3370629 GCGTGGTTCAGAGTCACTGGAGCGCGGCTTACCAGCGCGGCTGCTGGCTCTCTCG 3370570

QY 926 ACCAGGAGTGGCGCGCTCGAGCGCGGCTCGACCATCGAGCGGTCTCGGTGCGT 985
Db 3370569 AGCTGCGTGGAGAGTCACTAACCGCGCTGCGCGCTGAGTCAAGGAGTCACTGCGCGC 3370510

QY 986 TGTGAGCGCGCTGCGCTTATCCAGAGCGCGTGTATGCTTTCAACG---CGCGGCTA 1042
Db 3370509 ACTTCGACCATGCTTGGCGCGCTATTACCTGATCTGCGCTCGGAGGTTCGAGCAATC 3370450

QY 1043 TGGGAGTCCGCGGTGTGGGTACTTCCAGAGGGGCGTGGAGCTCAGCACCGCG 1102
Db 3370449 TGGCGGCTTCGACGCGATCGCTACGCGGTGCGGCTGCGGACGACGCAACCGCGAGCG 3370390

QY 1103 TCACGACGCGCGCTGAGTTCGACACACACCAAGGATCTGGGATCTGCTCCCGCTCA 1162
Db 3370389 CCGAGGAGTGTATGGCGATGACCGCGCGCGCTTTCGGGCCGAGGTCAAGCGCGCA 3370330

QY 1163 TGCTGTGATCGCGAGCACTCGCGGACGAATACCTCGGCATCTACTAGCGAGGCGC 1222
Db 3370329 TCATGATCGCACCTACGCGTTCGCGCGGCTACTACGAGCGCTATTACAACCAAGCGCG 3370270

QY 1223 AGAAGTGGCGGTGAGCTCGGCAAGCAGATCGACGCGCTCTCCAGGACCGGCTGCA 1282
Db 3370269 AGAAGTGGCGAGCTGATCGCGCGACCTCGACGCGGCTATCGTCTGCTGAGCTGC 3370210

QY 1283 TGCTGAACCGGACACCGCTTACCGTTGCGCAACGAGTGTGTAGCGGTGCGGCAAGACCA 1342
Db 3370209 TGGTGTGCGCGACGACCGCGACACCGCTTCCGATGAGAGGTGAGACATCCG 3370150

QY 1343 TGTCCATGATCCACGATGACGGGCAATCGATCTCTCAACAGTGGCGGCTGACCTCA 1402
Db 3370149 TGGCGATGTACTTTTTCGA-----CCTGTGACGCTGCGCGCTAACTGG 3370105

QY 1403 CCGTCAACCGGCGCTGACGCTGCCACCGG---TGCGGCGGAGAGGGCTCCCGGTTG 1459
Db 3370104 CCGGCCACTGCGGATGCTGTGCGCTGCGGCTCTCCCGGACGAGGTTGCGGTTG 3370045

QY 1460 GCCTCGAGTGAAGCGCGCACTTCGAGAGTTCGAGCTTACCGCACCGCGCGCGTGA 1519
Db 3370044 GCCTACAGATCATGGCGCGCGCATTTGGCCGACGACCGGCTCTACCGGCTGGCGCGGCTT 3369985

QY 1520 TCAGGCGCGCGCGC 1533
Db 3369984 ATGAGCGCGCGCGC 3369971

RESULT 7
US-07-612-673-1
; Sequence 1, Application US/07612673
; Patent No. 5260208
; GENERAL INFORMATION:
; APPLICANT: Petre, Dominique
; APPLICANT: Cerbelaud, Edith
; APPLICANT: Mayaux, Jean-Francois
; APPLICANT: Yeh, Patrice
; TITLE OF INVENTION: NOVEL POLYPEPTIDES, THE DNA SEQUENCES
; TITLE OF INVENTION: ALLOWING THEIR EXPRESSION, METHOD OF PREPARATION, AND
; TITLE OF INVENTION: UTILIZATION
; NUMBER OF SEQUENCES: 17
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Finnegan, Henderson, Farabow, Garrett &
; ADDRESSEE: Dunner
; STREET: 1300 I Street, N.W.
; CITY: Washington, D.C.
; COUNTRY: USA
; ZIP: 20005
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/07/612,673
; FILING DATE: 19901114
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: Potter, Jane E.R.
; REGISTRATION NUMBER: 33,332
; REFERENCE/DOCKET NUMBER: 03715.0010
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (202) 408-4000
; TELEFAX: (202) 408-4400
; INFORMATION FOR SEQ ID NO: 1:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1878 base pairs
; TYPE: NUCLEIC ACID
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: DNA (genomic)
US-07-612-673-1

Query Match 10.7%; Score 168.8; DB 1; Length 1878;
Best Local Similarity 48.0%; Pred.No. 4.4e-24;
Matches 703; Conservative 0; Mismatches 742; Indels 19; Gaps 7;
QY 66 CGCGCGCCACCATTCGACCTCGAGGAGTCCGCGCCAGCTCGTTCGGTCTGGC 125
Db 286 CGAAGGATTCAGGCATCACTTCGAAACAGCCGGCTCGAGTGGCGGCACTAT 345

QY 126 GGAGATGTGATCGCTTCGACCTGATCGAGAACTACCGCAACCGCGAGCCGCGAC 185
Db 346 CGACGAGCACTGGGCTCTACGACGCTGTCAGCAGTTGTACGCCAGAGGGACCCC 405

QY 186 GCGGTACAGGACCGGACATCGCGCGGAAACCGACCGGCGGACGAGACCGTTCAAGC 245
Db 406 GCGGACCACTGACGCGAGCAG---CGTCCCAAGTGCAGCGGAAATCTTTGAGCGC 462

QY 246 ATTCATCCGTTCTTCGCGGTGGAGGGCCACAGAGGGGCCACTGAGCGACCTGACCGC 305
Db 463 TTGGTATGTGACCAACCACTCCCGCGACGCTCGGACGGCGCTCTGACCGCGCGCGT 522

QY 306 CCGCATCAAGGACTGATCGCGCATCGCGGTATGCGCCACCGACGAGCGGTCCCGATCT 365
Db 523 GCGCATCAAGGACACGTCGACCGGTGGCGGAGTTCCGATGATGAACCGATCTCGACCGT 582

186 GCGGTACACGACCGGACATCGCGCGCAACCGACCGCGGACGAAGACCGGTTCAAGC 245
Db
406 GCGGACCACTGTCACGCGACGAG---CGGTGCAAGTGGAGCGAAGATCTTTGAGCG 462
Qy
246 ATTATCCGGTTCTGCGGGTGGAGCGGCCACGAGAGGGGCCACTGAGGACCTGACCGC 305
Db
463 TTGGTATGACACACAGATCCCGCGGACGTCGACGCGCTCTTACGCGCGGCGGT 522
Qy
306 CGGATCAAGGACTGATCGCCATCGCCGGTATGCCACACGAACGGGTCCCGGATGCT 365
Db
523 GCGGATCAGGACAAAGTACCGTGGCGGAGTTCCGATGATGAACGGATCTCGACGGT 582
Qy
366 CCGGACTGTATGCGCACAGAGATCCGTTGGTGGAGCGGCTGCTCGCGGACGGCGC 425
Db
583 AGAGGGAATTTACTCCGTCACGCGACGCACTGTGGTCACTCGACTACTGGCGCGGTGC 642
Qy
426 CACCATCGTCGCAAGCAACCTCGAGGACATGGGAT---GGGTATCGGTGAAGGAG 482
Db
643 AACCGTCGGGGAAGCTGTGTGAGAACCTGTGTTCTCGGTTGAGCTTCACACC 702
Qy
483 CGTCTACGCTCTGCGCTGAACCCGAACAAACCCCGCCACGGACGGGTGGATCTTCCAG 542
Db
703 GGAAGCGGACCGGTCCGCAATCCATGGACCGGACGGCAAGCAGGTGGATCATCCGG 762
Qy
543 CGGCTCGGGGTCGCTGCGTCCGCGCATGTGCACTTGCCTGGCGTGGGTGATGAGCG 602
Db
763 CGGAGTCAGCACTCGTCGAACCGTGAAGTGTGCAATTTTGCATCGCGGGATCAAG 822
Qy
603 AGGACATCCGATCCGCGCGCATGTCGCGACTGTCGCGCATGAAGCGCACCCACGG 662
Db
823 CGGATCGATCCGATCCCGCGCACTTTCGCGCGTGTGCGGCAAAAGCCAGCTTGG 882
Qy
663 CTGTTGCGCTCTTACGGCTGACATACATGACACACCTTGGACCATCGGGCCCAT 722
Db
883 GCTCGTCCGATACCGGTCATTTCCCATCGAGCGAACAATCGACCATCTCGGCCGAT 942
Qy
723 CACGAGGGGTGAGCTCAACCCCGGGTCTCTGAGGTGTTGGCGGGCGGAGTGGCG 782
Db
943 CACACGACGGTCCACGATGACGACTGATGCTCTCGGTATCGCGCGCGGACGGTAA 1002
Qy
783 CGACCTCAGTGGGTGATCTTCGGAGCGGAGAACTACGGTTCGCGGTGGGCA 842
Db
1003 CGACCCAC---GCCAAGCGCACAGTGTGGAAGCAGGTGATCTGTCCACCTCGACTC 1059
Qy
843 GCGAGTATCCGCTGAGATTCGCGTGTGCGAGGAGTCACTGAGCGCAACGGTCCGAC 902
Db
1060 CGATGTGACGGCTGCGAATCGAATCGTTTCGAGAGGATCCGCGACGCGCTCTACA 1119
Qy
903 GCGGAGGTGATCCCGCGTTCAACGAGGACTGCGGGGCTCGAGAGCGCGGTGCGAC 962
Db
1120 GCGGAGGTGAGACGCACTCGCGCGGACACAGTCTGACCGAATCGGTTGCAC 1179
Qy
963 CATCGAGGCTCGGTCGCTGTTGAGCGGCGCTGCGCTATCCAGAGG---CGGT 1019
Db
1180 GGTAGAGAGTAACATCCGTTGGATCTGATGCTTTTCCATCTGGAACGTGATCGC 1239
Qy
1020 GATGGCTTTCAACGCGCGCTATGCGGACTCCGCGGTGTGGGTACTTTCACAAAGG 1079
Db
1240 CACGAGCGTGGTGCCTACAGATGTGAGCGCAACGAGATACGGCATGAACGCGAAG 1299
Qy
1080 GCGGTGAGCTCAGCACCGCTCAGACGCGGCGGCGAGTTCGACCAACCAACGAAGA 1139
Db
1300 TTGTAGATCCGGAATGATGACACTTTGCTTCGACGATTCAGCACCGCCGACGC 1359
Qy
1140 TCTGGGATCTGTCGCGGTGATGCTGATCGGAGACCTCGCGGACGAATACCT 1199
Db
1360 TCTGTCGGAACCGTCAACTGGTGGCTTACCGGCGCACCGGATCACACCTCGG 1419
Qy
1200 CGGATCCATACGCGAGCGGACAGACTCGCGGTGAGCTCGGCAAGCAGATGACGC 1259
Db
1420 CGGCGGAGTACGGCAAGCCGGAACCTCGTACCGCTTTCGCGCGCGCTTACGACAC 1479
Qy
1260 CGTCTCCAGACCGGGCTGACTGTGACCCCGGACCAACGCTTACCGTTGCCAACGAGCT 1319

1480 TGCCTTGAGACAATTCGACTCTGCTGTGTGATGCAACGCTCCCTACGATCCGAAT 1539
Qy
1320 GTTGAGCGGTGCGGACAGACACCATGTCCATGATCCCAAGATGACGGGCAATCGATCCT 1379
Db
1540 GCCGGGAAAGACGTAGATCGTCAACCTTCATCAC---CAAGCTCTCGGATGATCGC 1596
Qy
1380 CAACACGTGCGGTGGACCTCACCGGTACCCGGGCGCTGACGGTGCACCGGTGCGGG 1439
Db
1597 CAACACGGCACCATTCGACTGACCGGACATCGTCTCTGCTCGGTCGCGCGCGCTG-- 1654
Qy
1440 CGAAGAGGCTGCCGCTTGGCTTCCAAAGTATAGCGCGCACTTCGAGAGTGCACGCT 1499
Db
1655 -GTGAACGGGCGCTCGGTGCGAATGATGATCACCGGACAGACATTCGACATGCGACGC 1713
Qy
1500 CTACCGCACCGGCGCGCTGATCGAGGCGCGCGC 1533
Db
1714 TCTTGTGTCGACGCGCATTCGAACACCTCGC 1747

RESULT 9

US-09-252-991A-9752
; Sequence 9752, Application US/09252991A
; Patent No. 6551795
; GENERAL INFORMATION:
; APPLICANT: Marc J. Rubenfield et al.
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS
; FILE REFERENCE: 107196.136
; CURRENT APPLICATION NUMBER: US/09/252.991A
; CURRENT FILING DATE: 1999-02-18
; PRIOR APPLICATION NUMBER: US 60/074,788
; PRIOR FILING DATE: 1998-02-18
; PRIOR APPLICATION NUMBER: US 60/094,190
; PRIOR FILING DATE: 1998-07-27
; NUMBER OF SEQ ID NOS: 33142
; SEQ ID NO 9752
; LENGTH: 1467
; TYPE: DNA
; ORGANISM: Pseudomonas aeruginosa
US-09-252-991A-9752

Query Match 9.6%; Score 151; DB 4; Length 1467;

Best Local Similarity 49.0%; Pred. No. 1.1e-20;

Matches 598; Conservative 0; Mismatches 590; Indels 33; Gaps 6;

Qy 273 GCGCACGAGGCGCACCTGAGCGACCTGACCGCGCGATCAAGACTGCAATCGCATCGC 332
Db 201 CGGCGAGAACCGGCGCTCTCGGCGCGCGATCGCCACAGGACCTGTTGACCCCA 260
Qy 333 CGGTATGCCACCAACGAGCGGTCCCGGATGCTCCGACTGTGATCGCCACCGAGGATGC 392
Db 261 GGGGTACGACACAGCTGCGGTCGAGATGCTCGACAACTTCGCTCGCCCTACGACGC 320
Qy 393 CGTGTGTGAGCGGCTGCTCGCGGAGGCGGCCACCATCGTCGGCAGAGACCTCGA 452
Db 321 CACCGTGTGAGAGTCAACCGCGCGCGGCTTACCTCGCGAGCTGAACATGGA 380
Qy 453 GGACATGCGGATGGG---TATCGGTGAAGGAGCGGTCTACCGTCTCGCTGGAACCCGAA 509
Db 381 TGAATTGCGCATGGCTCGTCGAACACAGTCCAGCGCATACGCGCGGTGAAGAACCCCTG 440
Qy 510 CAAACCCCGCCACGCGACGCGGTGATCTTCCAGCGGTCCCGCGTCCGCTCGCTCGCGG 569
Db 441 GAGCCTCAGCGGTGCGCGGCGCTCTCTCGCGGCTTCCCGCGGCGAGTCCGCGCGCG 500
Qy 570 CATGTGTGACTTCGCGCTTGGCGTTCGATGAGCGAGGACATCCGGATCCCGCGCGCATG 629
Db 501 CTTGTGCGCGCGCCACCGGACCGGATACCGCGGCTCGATCCGCAACCGCGGCGCT 560
Qy 630 GTGCGGATGTGTGCGCATGAGGCGGACCCACGCGCTGTGTGCGCTTTACGCGCTGACATA 689
Db 561 GACCAACCTCAGCGGATCAAGCAACCTACGCGCGGCTTTCCGCTGGGCGCATGATCGC 620

QY 630 CATGACACACCTTGGACCACTATCGGCGCCATCACAGGGGGTTCAGCTCAACGCCCG 749
Db 621 CTACCTTCCAGCTTCGACCAAGGGCGCCCTGGCGCGCACACCGGAGACTCGCGCT 680
QY 750 GGTCTCAGAGTGTGGCGGGCGGAGCTGCGCGGACCTCAGTGGGTGCGTACCTTCC 809
Db 681 GATCTGGGGTGTGGCGGATTCGATCCGAGGACTCGACACGCT-----CGAACA 734
QY 810 GGAGCGGAGACTACGCTCGCGCTCGGCGAGGAGTATCCGCTGAGATTCGGCT 869
Db 735 GCGGTGAGACACTACCTTGGCGGCGCTCGAAGACCGCTGAGCGGCTCGGCATCGCCT 794
QY 870 CGTCAGAGTCACTGGAGCGGAAACGGTGCAGCGCGGACCTGATCGCGGTTCAACCA 929
Db 795 GCGCGGGAATCTTCGCGCGGCTCGACAGCGCATCGCGCGGCTCGGCTGCGC-- 852
QY 930 GGGACTGCGCGCTCGAGAGCGCGGTGCACCATCGAGCGGTCTCGGTGCGCTGTG 989
Db 853 -GTAGTCAGAGCTGAAGACGCTCGGCGCACCGTGAAGACATTTCCCTGCGCAACAT 911
QY 990 GACGCGGCTGCGCTATCCAGAGCGGGTGTATGGCTTTCAACGCGCGCTATGGCGGA 1049
Db 912 GCAGACGCCATCCAGCTACTAGTGTATCGCGCGCGCGAGGCTCTCAACCTGTC 971
QY 1050 CTCGCCCGTGTGGCTATCTCCACAGGGCGGTGACGTCAAGACCGCGCTCAACAC 1109
Db 972 GCGCTTCAGCGGCTGCTATGGCTATCGTTGCGACCGCGCGGAGAACCTGGAAGACCT 1031
QY 1110 GCGCGCCAGAGTGCACACCAAGGATCTGCGCATCTGCGGATCTGCGGCTGCTGGT 1169
Db 1032 GTACAGCGCTGCGCGGAGGCTTCGCGAGGAGTGAAGAACCGCATCTGCTGG 1091
QY 1170 GATCGCGAGACCTGCGCGAGAACTACCTCGGCATCACTACGAGAGGCGGAGAACCT 1229
Db 1092 CACCTACGACTCTCG--GCGGCTACTAGATGCTTATACCTGCGAGGCTCAGAAGAT 1148
QY 1230 GGGGTGAGCTCGCAAGCAGTACGCGCGTCTCCAGACCGGCTGCTGCTGAC 1289
Db 1149 TCGCGGTGATCAAGACGACTTCTGTCAGCGCTTCCGAGTGAAGTGAAGTCTCGG 1208
QY 1290 CCGGACACGCTACCTGTGCAACAGCTGTGAGCGGTGCGAAGACACCATGTCCAT 1349
Db 1209 CCGGACACGCGAACCGCGCTGGAAGATCGCGGAGAGAACGACGACCGGTTTCCCA 1268
QY 1350 GATCCACGAGTACGCGCAATGCGATCTCAACAGTGCCTGCGGTGAGCTCACCGTCA 1409
Db 1269 GTACCTG-----GAAGACATCTACACCATCACCGCAACCTCGCGGCT 1313
QY 1410 CCGCGGCTGACGCTGCGGCGGAGGAGGCGCTGCGGCTGCGGCTGCGGCTGCGGCT 1469
Db 1314 GCGGCGGTGTCATGCGCGCGGCTTCTGCG--ACGGCTGCGGCTGCGGCTTCTG 1370
QY 1470 GATAGCGGCACTTCGAGGA 1490
Db 1371 GCTCGCGCTACTTCCAGGA 1391

RESULT 10

US-09-252-991A-10084/c
; Sequence 10084, Application US/09252991A
; Patent No. 6551795
; GENERAL INFORMATION:
; APPLICANT: Marc J. Rubenfield et al.
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS
; TITLE OF INVENTION: AERUGINOSA FOR DIAGNOSTICS AND THERAPEUTICS
; FILE REFERENCE: 107196.136
; CURRENT APPLICATION NUMBER: US/09/252,991A
; CURRENT FILING DATE: 1999-02-18
; PRIOR APPLICATION NUMBER: US 60/074,788
; PRIOR FILING DATE: 1998-02-18
; PRIOR APPLICATION NUMBER: US 60/094,190
; PRIOR FILING DATE: 1998-07-27

; NUMBER OF SEQ ID NOS: 33142
; SEQ ID NO 10084
; LENGTH: 1818
; TYPE: DNA
; ORGANISM: Pseudomonas aeruginosa
US-09-252-991A-10084

Query Match 9.6%; Score 151; DB 4; Length 1818;

Best Local Similarity 49.0%; Pred. No. 1.1e-20;

Matches 598; Conservative 0; Mismatches 590; Indels 33; Gaps 6;

QY 273 CGCCACGAGGCGGCACCTGAGCGAGCTGACCCCGCGGATGCCACAGGAGCTGTCATCGCATCGC 332

Db 1612 CGCGAGAACGCGCCCTCTCGGCGCGGATGCCACAGGAGCTGTCATCGCATCGC 1553

QY 333 CGGTATGCCACACAGAACGGGTCCCGGATGTCGAGTGTGATGCGCACCGAGATGC 392

Db 1552 GGGCGTACGACACGAGTTCGAGATGTCGACAACTTCGTCCTGCCCTCAGAGC 1493

QY 393 CGTGTGTGTGAGCGGCTGTCGCGCGAGCGCCACCATCGTCGCGCAAGCAACCTCGA 452

Db 1492 CACCGTGTGAGAGAGCTCACCGCGCGCGCTTACCTCGGAGCTGAACATGA 1433

QY 453 GGNATGGCGATGGG---TATCGGTGAAGGAGAGCTTACGTCCTGCTGCTGAAACCGGAA 509

Db 1432 TGAATTCGCCATGGGCTCGTGAACCAAGTCCAGCCACTACGCGCGGTTGAAGAACCCCTG 1373

QY 510 CAAACCGCGCCACGCGCACGGGTGATTTCCAGCGGCTCCGCGCTGCGCTGCTGCGG 569

Db 1372 GAGCTCGACCGGTGCGGCGGCTCTCCGCGGTTCCGCGCGGAGTTCGCGCGG 1313

QY 570 CATGTGCTGATTCGCTGCGCTGCGGTGATGAGGAGGAGCATCGGATCCCGCGCGCATG 629

Db 1312 CTGTGTCGCGCGCGCACCGGACCGATACCGCGCGCTCGATCCGCAACCGCGCGCT 1253

QY 630 GTGCGGACTGTGCGGATGAAGCGGACCGACCGGCTGCTGCTTACGCGCTGACATA 689

Db 1252 GACCAACCTACCGGATCAAGCAACCTACGCGCGGTTCCGCTGGGGGATGATCGG 1193

QY 690 CATGACCAACACTTGGACCACTCGGCGCCATCCAGGGGGGTGAGCTCAAGCGCGG 749

Db 1192 CTACGCTTCAGCTCGACGCGCGCGGCTGCGCGGCTGCGCGGCTGCGCGGCT 1133

QY 750 GGTCTCGAGGTGTGGCGGCGGCGGCTGCGCGGCTGCGCGGCTGCGCGGCTGCGCGGCT 809

Db 1132 GATGCTGGGGGTGATGGCGGATTCGATCCGAGGACTCGACAGGCTCGACAGGCT 1079

QY 810 GGAGCGGAGAACTACGCGCTCGCGCTCGCGGAGGAGTATCCGCTCTGAGATTCGCGGT 869

Db 1078 GCGGTGAGGAGTACTGCGCGCTCGCGGCTGCGAGGAGGCTGAGCGGCTGCGGCT 1019

QY 870 CGTCGAGGAGTCACTGAGCGGAAACGGTTCGAGCGCGGAGGAGTATCGCGGCTTCAACA 929

Db 1018 GCGCGGGAATATCTCGCGCGGCTCGACAGCGGCTGCGCGGCTGCTGCGGCT 961

QY 930 GGGACTGCGCGGCTCGAGAGCGGCTCGAGGCTCGAGGCTCGAGGCTCGAGGCTCGAGGCT 989

Db 960 -GTAGTCGAGGAGCTGAAGACGCTCGGCGCCAGCTGAGGAGCATTTCCCTGCCGAACAT 902

QY 990 GACGCGGCTGCGCTATCAAGCGGCTGATGCTTCAAGCGCGGCTGATGCGGAG 1049

Db 901 GCAGCACGCTTCCAGGCTTCTGATGATCGCGCGGCGGCTTCTTCAACCTGTC 842

QY 1050 CTCGCGGCTGCGGCTTCTTCAAGGGGCGGCTGAGAGCTGAGCGCGGCTGAGCGGCT 1109

Db 841 GCGCTTCGAGCGGCTGCGCTATGGCTATGCTTGGAGCGGCTGCGGAGCTGAGAGCT 782

QY 1110 GCGGCGGAGAGTTCGACCAACCGGATCTGGGATCTGCTGCTGCTGCTGCTGCTGCT 1169

Db 781 GTACAGCGCTCGCGCGGAGGCTTTCGCGAGCGGAGTGAAGAAACCGCATCATGCTCGG 722

QY 1170 GATCGGAGGAGCTGCGGAGGAGATACCTCGGCATCCACTACGCGGAGGAGGAGGAGCT 1229

Db 721 CACTACCGACTCTCG---GCCGGCTACTAGATGCGCTATTACTCGAGGCTCAGAGAT 665
QY 1230 GCGCTGGAGCTCGGCAAGCAGATCGACGCGCTCTCCAGGACCGGGCTGCACTGCTGAC 1289
Db 664 TCGCGGCTGATCAAGAACTCTGTCAGCGCTTTGCCGAAGTGAAGTATCTCTCGG 605
QY 1290 CCGACACCGCTACGTTGCCAACGAGCTTTGAGGGTTCGCGAGACACCATGTCAT 1349
Db 604 CCGACACCGCTACGTTGCCAACGAGCTTTGAGGGTTCGCGAGACACCATGTCAT 545
QY 1350 GATCCCGGATGACGGGCAATCGATCTCTCAACAGCTGCGCCCTCGACCTCACCGGTCA 1409
Db 544 GTACCTG-----GAAGACATCTACACCATCACCGCAACCTCGCGGCT 500
QY 1410 CCGCGGCTACGCTGCCACGGTGGCGGAGAGGGGCTCGCGCTTCCCAAGT 1469
Db 499 GCGGGGCTGTCCATGCGCCCGCTTCGTGCG---ACGGCTGCGGGTGGTGTCCAGTT 443
QY 1470 GATAGCGCGCACTTCAGGA 1490
Db 442 GCTCGCGCTACTTCCAGGA 422

RESULT 11

US-09-103-840A-2
; Sequence 2, Application US/09103840A
; Patent No. 6294328
; GENERAL INFORMATION:
; APPLICANT: FLEISCHMAN, Robert D.
; APPLICANT: WHITE, Owen R.
; APPLICANT: FRASER, Claire M.
; APPLICANT: VENTER, John C.
; TITLE OF INVENTION: DNA SEQUENCES FOR STRAIN ANALYSIS IN MYCOBACTERIUM
; FILE REFERENCE: 24366-20007.00
; CURRENT APPLICATION NUMBER: US/09/103,840A
; CURRENT FILING DATE: 1998-06-24
; NUMBER OF SEQ ID NOS: 2
; SOFTWARE: Patent In Ver. 2.1
; SEQ ID NO 2
; TYPE: DNA
; ORGANISM: Mycobacterium tuberculosis
; FEATURE:
; OTHER INFORMATION: CDC 1551
; OTHER INFORMATION: "n" bases at various positions throughout the sequence
; OTHER INFORMATION: represent a, t, c or g
US-09-103-840A-2

Query Match 7.8%; Score 123.4; DB 3; Length 4403765;
Best Local Similarity 53.2%; Pred. No. 4.6e-15;
Matches 309; Conservative 0; Mismatches 266; Indels 6; Gaps 2;
QY 262 CCGGTGGAGGGCGCCACGAGGGGCCACTGAGCAGCTGACCGCGGATCAAGGACTGC 321
Db 2641838 CCGGGGCGCGCGGACACGCGCGCTGTGGGCACTTCGATCGCGTCAAGGACGAC 2641897
QY 322 ATCGCCATCGCGGTATGCCACACGAAACGGGTCCCGGATGCTCCGACTGTGATCGCC 381
Db 2641898 GTCGACGTTGTGGAGTGCACACCGCTTCGGCACCCAGGG---CTATGTGCGCGCTGCT 2641954
QY 382 ACCGAGGATGCGTGGTGGTGGAGCGGCTGTGCGCGGAGGCGGCACCATGTCGGCAAG 441
Db 2641955 ACCGAGACTGTGAGTGTGTCGGGCGCTCAAGGCGCGGAGCGGTGATGTCGGCAAG 2642014
QY 442 AGAACCCTCGAGGACATGG---CGATGGGTATCGGTGAAGCGAGCGTCTACGGTCTTGGG 498
Db 2642015 ACGAATCTTGTGAATTGGGCGAGTGGCGCTTACACGAGACCGGGTTCGGACACACC 2642074
QY 499 CTGAACCCGAAACACCCCGCCACGCGGCTGTGATCTTCAGCGGCTCGGCGCTGCC 558
Db 2642075 CGCAACCCCTGTGCGGCGGACACGCGGGTGTGATCTCTGGGCGGTAGCGCGGGCG 2642134

QY 559 GTCGTGCGCGCATGCTCGACTTCCCTCGCTGGCGTTCGATGAGCGAGCAGCATCGGATC 618
Db 2642135 GTGGCGCGCGCTGGTTTACCGCGCTATCGGCTCCGACGGCGCGGAGCATCCGATC 2642194
QY 619 CCGSCCGCATGCTGCGGACTGCTCGGCATGAAGCGGACCCACGGCCTGCTGTCCTTAC 678
Db 2642195 CCGCGAGATGACACACACTAGTGGGCATCAAGCACAAACGCGTCTGATCTCCACTGG 2642254
QY 679 GGCCTGATATATGAGACACACACTTGGACACATCGGGGCCCATCACAGGGGGTCTGAG 738
Db 2642255 CCGCTGCCGAGGCGGTTCAACGCGCTCACGGTCAACGCGCTACTGGGCCGCACTGTGAG 2642314
QY 739 CTCACGCGCGGCTCTCTGAGGTGTTGGCGGGCGGCGGCGGCGGCGGCGGCGGCGG 798
Db 2642315 GATGGGCGCTGCTGCTGAGCGCGGCTCCGCAAGCTCGAGGGCGGACCGCCACACCA 2642374
QY 799 CGTAACCTTCGCGAGCGAGAACTACGCGCTCGGGCTCGG 839
Db 2642375 CCCCCGCTGACGGTGTCCGATTTCGTGCGCATCGCCCTGG 2642415

RESULT 12

US-09-103-840A-1
; Sequence 1, Application US/09103840A
; Patent No. 6294328
; GENERAL INFORMATION:
; APPLICANT: FLEISCHMAN, Robert D.
; APPLICANT: WHITE, Owen R.
; APPLICANT: FRASER, Claire M.
; APPLICANT: VENTER, John C.
; TITLE OF INVENTION: DNA SEQUENCES FOR STRAIN ANALYSIS IN MYCOBACTERIUM
; FILE REFERENCE: 24366-20007.00
; CURRENT APPLICATION NUMBER: US/09/103,840A
; CURRENT FILING DATE: 1998-06-24
; NUMBER OF SEQ ID NOS: 2
; SOFTWARE: Patent In Ver. 2.1
; SEQ ID NO 1
; LENGTH: 4411529
; TYPE: DNA
; ORGANISM: Mycobacterium tuberculosis
; OTHER INFORMATION: H37Rv
US-09-103-840A-1

Query Match 7.8%; Score 123.4; DB 3; Length 4411529;
Best Local Similarity 53.2%; Pred. No. 4.6e-15;
Matches 309; Conservative 0; Mismatches 266; Indels 6; Gaps 2;
QY 262 CCGGTGGAGGGCGCCACGAGGGGCCACTGAGCAGCTGACCGCGGATCAAGGACTGC 321
Db 2644543 CCGCGGCGCGCGGACACGCGCGCTGTGGGCACTTCGATCGCGTCAAGGACGAC 2644602
QY 322 ATCGCCATCGCGGTATGCCACACGAAACGGGTCCCGATGCTCCGACTGTGATCGCC 381
Db 2644603 GTCGACGTTGTGGAGTGCACACCGCTTCGGCACCCAGGG---CTATGTGCGCGCTGCT 2644659
QY 382 ACCGAGGATGCGTGGTGGTGGAGCGGCTGTGCGGCGAGGCGGCACCATGTCGGCAAG 441
Db 2644660 ACCGAGACTGTGAGTGTGTCGGGCGCTCAAGGCGCGGAGCGGTGATGTCGGCAAG 2644719
QY 442 AGAACCCTCGAGGACATGG---CGATGGGTATCGGTGAAGCGAGCGTCTACGGTCTTGGG 498
Db 2644720 ACGAATCTTGTGAATTGGGCGAGTGGCGCTTCAACGAGGAGCCCGGTTTCGGACACACC 2644779
QY 499 CTGAACCCGAAACACCCCGCCACGCGGCTGTGATCTTCAGCGGCTCCGCGCTGCG 558
Db 2644780 CGCAACCCCTGTGTCGCGCGGACACGCGGGTGTGATCTCTGGGCGGTAGCGGGGGCG 2644839
QY 559 GTCGTGCGCGCATGCTCGACTTTCGCCCTGGGCGTTCGATGAGCGAGGACATCCGATC 618
Db 2644840 GTGCGCGCGGCTGTGTACCGCGCTATCGGCTCCGACGCGCGCGGAGCATCCGATC 2644899
QY 619 CCGCGCGCATGCTGCGGACTGTGTCGCGATGAAGCGGACCCACGCGCTGTGTCGCTTAC 678

Db 2644900 CCGGACGATGGACACACACCTAGTGGGCATCAAGCCACACGCGGTGGATCTCCACCTGG 2644959
QY 679 GGCCTGACATACATGACACACACCTTGGACACATCGGGCCCATCACAGGGGGGTGAG 738
Db 2644960 CCGTCCGGAGGCGTTCAACGGCGTCAAGGTCAACGGGGTACTGGCCCGCACTGTGGAG 2645019
QY 739 CTCACGCCCGGTCTCTGAGGTGTGGCGGGGGCGGACTGCGGCAACCTCAGTGGGTG 798
Db 2645020 GATCGGGCGCTGGTGTGACCGCGCTCCGGCAACGTCGAGGGCGCGCCACAGCCA 2645079
QY 799 CGTAACCTTCGGAGCGCGGAGAACTACGGCTCCGGGCTCGG 839
Db 2645080 CCCCCGTCACGGTGTCCGATTTCGTGCGGCACTGCCCCCTGG 2645120

RESULT 13

US-08-233-146-8
; Sequence 8, Application US/08233146
; Patent No. 5648256
; GENERAL INFORMATION:
; APPLICANT: BEPPU, TERUHIKO
; APPLICANT: YAMADA, HIDEAKI
; APPLICANT: NAGASAWA, TORU
; APPLICANT: HORINOUCHI, SUEHARU
; APPLICANT: NISHIYAMA, MAKOTO
; TITLE OF INVENTION: A GENE ENCODING A POLYPEPTIDE HAVING
; TITLE OF INVENTION: NITRILE HYDRATASE ACTIVITY, A TRANSFORMANT CONTAINING THE
; TITLE OF INVENTION: GENE AND A PROCESS FOR THE PRODUCTION OF AMIDES USING THE
; TITLE OF INVENTION: TRANSFORMANT
; NUMBER OF SEQUENCES: 8
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: PENNIE & EDMONDS
; STREET: 1155 AVENUE OF THE AMERICAS
; CITY: NEW YORK
; STATE: N.Y.
; COUNTRY: U.S.A.
; ZIP: 10036
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/233,146
; FILING DATE: 22-APR-1994
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/694,746
; FILING DATE: 02-MAY-1991
; ATTORNEY/AGENT INFORMATION:
; NAME: MISROCK, S. LESLIE
; REGISTRATION NUMBER: 18,872
; REFERENCE/DOCKET NUMBER: 7005-024-999
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (212) 790-9090
; TELEFAX: (212) 869-9741
; INFORMATION FOR SEQ ID NO: 8:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 2456 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: DNA (genomic)
; ORIGINAL SOURCE:
; ORGANISM: Pseudomonas chlororaphis
; STRAIN: B23 (FERM BP-187)
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 1021..1620
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 1666..2325

US-08-233-146-8
Query Match 7.6%; Score 120; DB 1; Length 2456;
Best Local Similarity 47.9%; Pred. No. 1e-14;
Matches 444; Conservative 0; Mismatches 470; Indels 12; Gaps 3;
QY 614 GGATCCCGCGCGATGGTGGGACTGGTGGGATGAAGCGACCCACGCGCTGGTGGCGGT 673
Db 1 GGATCCCGCTGGCGCTTCTGGGTACTTACGGCATGAAGCCACCCACGCGCTGGTGGCCT 60
QY 674 CTTACGGCCTTGACATACATGAGACCAACACCTTGGACCACATCGGGGCCCATCACAGGGGG 733
Db 61 ACACCGCGCTCATGGCGATTGAAGCACGATCATCATGTGCGGCCCATCACCGGTAACG 120
QY 734 TCGAGCTCAAGCGCCGGTCTCTGAGGTGTTCGCGGGCCGAGCTGGCGGAGCCTCAGT 793
Db 121 TCGCGGACAAACGGCGCTGATCTCGAGCAATGGCCGGTGCAGACGACTCGACCCCGGCC 180
QY 794 GGGTGGCTAACTTCGGAGCCGAGAACTACCGGCTCCGGCTCGCGAGGGAGTATCCG 853
Db 181 AGCG-----GGCGCTCAGTCTGATGACTATTTCAGTTACTTGAAGAGCGGTGAGCG 234
QY 854 GTCTGAGATTGGCGGTCTGCGAGGAGTCACTGGAGCGGAACGGTGCAGCAGCGGACGTGA 913
Db 235 GACTCAGATCGGGGTGTTGCAAGAGGGATTTCGGCTTGCTAACACGAGACCCCTCGCGTGG 294
QY 914 TCGCGCGGTTCAACACGAGGACTGGCGCGCTCGAGAGCGCGGTGCGACCATCGAGCGGG 973
Db 295 CGGCAAGTGGCGGAGCGCATCGCCGACTCGAGGGTGGGCGCTCATGTGAGCGCG 354
QY 974 TCTCGGTGCGGTGTGGAACCGCGCGCTGGGCTTATTCAGAGCGCGGTGATGGCTTTCAACG 1033
Db 355 TCTCATTCGCGAGCAACACTGGCAGGGTGTGTGGCACCCCATCGGTTGGAGAGGCT 414
QY 1034 CGCGCGCTATGGCGGACTCGCGCGGTGGGCTTCCACAAAGGGCGGTGAGCGTCA 1093
Db 415 TGACCATGAGATGATGATGCAATGGCAACGGCGCAGGCTTTAATCGAAGAGACTTTACGATG 474
QY 1094 GCACCGCGCTCAGCAGCGCGCCAGAGTGCACACCCACACAGGATCTGGCGATCTGT 1153
Db 475 TCGGCTGTGGACAAACAGCAGCTGGCGCGACGACGACGACCAATATTCGCGGTGCG 534
QY 1154 CCGGCTGATGCTGGTGTGATCGCGGACGACCTGGCGGCAATACCTCGGCTACCTACG 1213
Db 535 TCAAGCTCTGCTGCTGCTGGCGCAATACGCGCTGTGCGCTTACAGCGAGCGTACTACG 594
QY 1214 CGAAGCGCGAGAACCTCGCGCTGGAGCTCGGCAAGCAGATCGACGCGCTCTCCAGGACC 1273
Db 595 CCAAGGCCAGAACCTTGCACGCTTTCGCGCGAGGATACGACAAAGCGCTGCACACCT 654
QY 1274 GGGCTGCACTGTGACCCCGACACGCTACCGTTGCCACAGAGCTGTGGAGGGTGGC 1333
Db 655 ATGACCTGTGGTGTGATGCCACCGCCCATCGGCGCCCAACCCCGCCAG---CGA 711
QY 1334 AAGACACCATGTCCATGATCCACGATGACGGGCAATGCGATCTCAACACGCTGCCCGC 1393
Db 712 ACTGCTGATCAGGAGTACGTGGCTCGCGGTGGAAATGATCGGCAATACCGGCGCAC 771
QY 1394 TGGACCTCAGCGTCAACCGCGCTGACGCTGCCCGCGGTCAGGTCGGGGCAGAGGGCTGCG 1453
Db 772 AGGACATCACCGGGCATCCGGCCATGTCGATTTC---CGTGTGGCTGCTGAGAGGGCTGCG 828
QY 1454 CCGTTGGCTCCAAGTATAGGCGCCACTTCGAGGAGTGCAGCGCTCTACCGACCGCGCG 1513
Db 829 CCGTCGGGCTGATGCTGTCGAAAACACTACGCGGAGGCGACGATTACCAAGCGCGG 888
QY 1514 CCGTATCGAGGCGCGCGGCTATGG 1539
Db 889 CGGGGTTGAAGCCTCGGTGGACTGG 914

RESULT 14

US-08-463-470-8

Sequence 8, Application US/08463470
Patent No. 5789211
GENERAL INFORMATION:
APPLICANT: BEPPU, TERUHIKO
APPLICANT: YAMADA, HIDEAKI
APPLICANT: NAGASAWA, TORU
APPLICANT: HORINOBUCHI, SUHARU
APPLICANT: NISHIYAMA, MAKOTO
TITLE OF INVENTION: A GENE ENCODING A POLYPEPTIDE HAVING
NITRILE HYDRATASE ACTIVITY, A TRANSFORMANT CONTAINING THE
TITLE OF INVENTION: GENE AND A PROCESS FOR THE PRODUCTION OF AMIDES USING THE
TITLE OF INVENTION: TRANSFORMANT
NUMBER OF SEQUENCES: 8
CORRESPONDENCE ADDRESS:
ADDRESSEE: PENNIE & EDMONDS
STREET: 1155 AVENUE OF THE AMERICAS
CITY: NEW YORK
STATE: N.Y.
COUNTRY: U.S.A.
ZIP: 10036
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent In Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/463,470
FILING DATE: 05-JUN-1995
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/233,146
FILING DATE: 22-APR-1994
APPLICATION NUMBER: US 07/694,746
FILING DATE: 02-MAY-1991
ATTORNEY/AGENT INFORMATION:
NAME: MISROCK, S. LESLIE
REGISTRATION NUMBER: 18,872
REFERENCE/DOCKET NUMBER: 7005-024-999
TELECOMMUNICATION INFORMATION:
TELEPHONE: (212) 790-9090
TELEFAX: (212) 869-9741
INFORMATION FOR SEQ ID NO: 8:
SEQUENCE CHARACTERISTICS:
LENGTH: 2456 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: DNA (genomic)
ORIGINAL SOURCE:
ORGANISM: Pseudomonas chlororaphis
STRAIN: B23 (FERM BP-187)
FEATURE:
NAME/KEY: CDS
LOCATION: 1021..1620
FEATURE:
NAME/KEY: CDS
LOCATION: 1666..2325
US-08-463-470-8

Query Match
Best Local Similarity 7.6%; Score 120; DB 1; Length 2456;
Matches 444; Conservative 0; Mismatches 470; Indels 12; Gaps 3;

QY 614 GGATCCCGCGCGATGGTGGGACTGGTGGGATGAAGCGACCCACGCGCTGGTGGCGGT 673
DB 1 GGATCCCGTGGCGCTTCTGGGTACCTAGGATGAAGCGACCCACGCGCTGGTGGCGGT 60
QY 674 CTTACGGCTGACATACATGACACCACTTGGACACATCGGGGCCATCACAGGGGG 733
DB 61 ACACCGGGCTGTGGGATGAAGCGACCATGATCATGTGCGGCCCATCACCGGTAACG 120
QY 734 TCGAGCTACAGCCCGGCTCTGAGGTGTGGCGGGCGCGACTGGCGGACCTCAGT 793

Db 121 TCGCGACAAACGCGCTGATGTGACGGCAATGGCCGCTGCAGACGAGTCTCGACCCGCGCC 180
QY 794 GGGTGGTAACTCTCGGAGCCGGAGAACTACGGCTCGCGCTCGCGAGGAGTATCG 853
Db 181 AGCG-----GCGCTCAGGTGATGACTATTGAGTTTACCTGGAAGGCGTGGAGCG 234
QY 854 GTCTGAGATTCCGGTCTGCGAGGATCACTGAGACCGGAAACCGTGCAGCCGCGAGTGA 913
Db 235 GACTCAGATCGGGTGTGCAAGAGGATTCGCGCTTCTAACCAGGACCTCTCGGTGG 294
QY 914 TCGCCGCTTAAACAGGACTGGCGGCTCGAGAGCCGCTGGAGCCGCTGAGCGGG 973
Db 295 CGACAAAGTGGCGACGCCATCGCCGCTCGAGCGTGGCGCTCATGTGAGCGCG 354
QY 974 TCTCGGTGCGGTGTGACGCGGCGCTGCGCTTATCCAGAGCGCGTGTGCTTCAACG 1033
Db 355 TCTCATTCGCGAGCACACTGGCAGGGTGTGTGGCACCCCATCGGTTCGAAGGCT 414
QY 1034 CGCGCTATGGCGACTCGCGGCTGGGGTCTATTCCAGAGGGCGCTGGAGTCA 1093
Db 415 TGACCATGCGATGATGATGCAACGCGCGAGGCTTTAACTGGAAGACTTTACGATG 474
QY 1094 GCACCGCGCTACGACGCGGCGCCAGAGTCGACACCCACCAAGGATCTGGCGATCCTGT 1153
Db 475 TCGGCTGTGTCGACAAAGCAGCTGGCGGACGACGACACCAATATTCGCGTCGC 534
QY 1154 CCGGCTGTGTCGAGTGGGAGCACTGCGGACGAAATACCTCGGATTCATACG 1213
Db 535 TCAAGCTCTGATGTCGCGCAATACGCGCTGTGCGGCTACACGCGACGCTACTAG 594
QY 1214 CGAAGGCGCAGACCTCGGCTGGAGCTCGGCAAGCAGATCGACGCGCTCTCCAGACC 1273
Db 595 CCAAGGCCAGAACCTTGCAGCTTTCGCGGAGGATACGACAAAGCGCTGCAACCT 654
QY 1274 GGGTGCATGTGACCCCGACCAACCGCTTACCGTTGCCAACGAGCTGTGAGCGGTGGC 1333
Db 655 ATGACCTGTGTGATGCGCACCCACCGCCCATCACGSCCAACCCACCGCCAG---CGA 711
QY 1334 AAGACACCATGTCCATGATCCACGATGACGGCAATCGATCTCTCAACAGTGCCTGC 1393
Db 712 ACTGCTGATCAGGAGTACGTGGCTGCGCGCTTGGAAATGATCGGCAATACCGCGCAC 771
QY 1394 TGGACCTCACCGGTACCCCGCGCTGACGCTGCCACGCGTGGCGGCGGAGAGGCTGC 1453
Db 772 AGGACATCACCGGCGATCCGCGCATGTCGATT---CGTGTGCGCTGTGGAGCGCTGC 828
QY 1454 CGTTGGCTTCAAGTATAGGCGCGCATTCGAGGATCGAGCTCTACCGCACCGGCG 1513
Db 829 CCGTCGGGCTGATGCTGTGCGCAAAACATACACCGGAGGCGACGATTTACCAAGCGCG 888
QY 1514 CCGTGTGCGGCGCGCGCGCTATGG 1539
Db 889 CGCGTTTGAAGCTCGTGGACTGG 914

RESULT 15
US-09-252-991A-15844
Sequence 15844, Application US/09252991A
Patent No. 6551795
GENERAL INFORMATION:
APPLICANT: Marc J. Rubenfield et al.
TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS
TITLE OF INVENTION: AERUGINOSA FOR DIAGNOSTICS AND THERAPEUTICS
FILE REFERENCE: 107196.136
CURRENT APPLICATION NUMBER: US/09/252,991A
CURRENT FILING DATE: 1999-02-18
PRIOR APPLICATION NUMBER: US 60/074,788
PRIOR FILING DATE: 1998-02-18
PRIOR APPLICATION NUMBER: US 60/094,190
PRIOR FILING DATE: 1998-07-27
NUMBER OF SEQ ID NOS: 33142
SEQ ID NO 15844
LENGTH: 1335

TYPE: DNA
; ORGANISM: Pseudomonas aeruginosa
US-09-252-991A-15844

Query Match 7.3%; Score 114.6; DB 4; Length 1335;
Best Local Similarity 52.7%; Pred. No. 1.1e-13;
Matches 280; Conservative 0; Mismatches 239; Indels 12; Gaps 1;

QY	258	CTGCCGGGTGGAGGGCCACGAGGGGCGCACTGAGCGACCTGACCGCGCGGATCAAGGA	317
Db	154	CGCCCGGGCGGTGGAGCGCCCTCGGTCCGTGGACGGCATTCCTACACCGCAAGGA	213
QY	318	CTGCATCGCCATCGCGGTATGCCACACGAGCGGTCCCGATGCTCCGACTGTGAT	377
Db	214	CAGTACCTGTGTAAGGGCTTACCGCGCTTCGCGAGCCCGGGTTCAAGACCTGGT	273
QY	378	CGCCACCGAGGATGCGTGGTGGAGCGGTGTTCGCGCAGCGCCACCATCGTCGG	437
Db	274	GGCCACGCGATGCTTACCGTTCGAGCGCTTCGCGCGCGCGGGGATCTGCTGGG	333
QY	438	CAGACGACCTCGAGGACATGGCATGGGTATCGGTGAAGGAGCGGTCTAGGTCCTGC	497
Db	334	CAGACCAACATGCCGCCCATGSCCAACCGCGGCATGACGCGGCGTCTACGGCCGCGC	393
QY	498	GCTGAACCCGAAACACCCGCCACCGCACGGG-----TGGATCTTCCAGCGG	545
Db	394	GGAGAGCCCGTACATATGCGCTTACCTCACCGCGCCCTTCGCTCGGGTCTCTCAATGG	453
QY	546	CTCCGGGCTGCGTCCGCTGCGGCGATGTCGACTTCGCCCTGGGGTCTGATGAGGAGG	605
Db	454	CGCCGACCGCCACCGCGGCCAGCTTCGCCCTTCGGCTTCGGCGAGGAAACCTGGTC	513
QY	606	CAGCATCCGATCCCGCGCGCATGTGTCGGAATGTCGATGAAGCGACCCACGGCT	665
Db	514	GACCGGGCGCGCGCGCTCGAACAACGGCTGTGCGCTACACCCCTTCGCGCGGGT	573
QY	666	GGTCCGCTTACGGCGCTGACATATGACACACCTTGGAGCACATCGGCCCATCAC	725
Db	574	GATCTCGGTGCGCGCAACTGGCCGCTGACCGCGACCATGAGCGTGGTGGTCCCTATGC	633
QY	726	CAGGGGGTTCGAGCTCAACCGCGGGTCTCTGAGGTGTTGGCGGGCGGA	776
Db	634	GCGGAGCATGGCGACCTGCTGGAATCTCTGACGTAGTGTGCGCGAGA	684

Search completed: June 27, 2004, 10:03:41
Job time : 168.584 secs

Result No.	Score	Query		DB	ID	Description
		Match	Length			
1	1578	100.0	1578	17	US-10-658-691-4	Sequence 4, Appli
2	208	13.2	1503	15	US-10-156-761-2732	Sequence 2732, Ap
3	208	13.2	9025608	15	US-10-156-761-1	Sequence 1, Appli
4	175.2	11.1	1479	13	US-10-282-122A-45553	Sequence 2553, A
5	172	10.9	1478	13	US-10-282-122A-26120	Sequence 26120, A
6	170.4	9.8	1485	13	US-10-282-122A-28642	Sequence 28642, A
7	152.2	9.6	1488	13	US-10-282-122A-14078	Sequence 14078, A
8	151	9.6	1455	9	US-09-815-942-7924	Sequence 7924, Ap
9	151	9.6	1455	13	US-10-282-122A-30488	Sequence 30488, A
10	142.6	9.0	1362	15	US-10-156-761-7248	Sequence 7248, Ap
11	141.4	9.0	1488	13	US-10-282-122A-11356	Sequence 11356, A
12	141.4	9.0	1536	13	US-10-282-122A-15096	Sequence 15096, A
13	140.6	8.9	1449	13	US-10-282-122A-31392	Sequence 31392, A
14	138.8	8.8	1512	15	US-10-156-761-7324	Sequence 7324, Ap

181 CGAGCCGCTACACGACCGGACATCGGCCCGCGAACCGACCGCGGACGAGACCGGTTTC 240
181 CGAGCCGCTACACGACCGGACATCGGCCCGCGAACCGACCGCGGACGAGACCGGTTTC 240
241 AAGCATTATCATCGGTTCTCGCGGTTGAGAGGCGGCGACCGAGAGGCGCACTGACGACCTG 300
241 AAGCATTATCATCGGTTCTCGCGGTTGAGAGGCGGCGACCGAGAGGCGCACTGACGACCTG 300
301 ACCGCGCGGATCAAGGACTGCAATCGGCATCGCGGTTATGCCACACGAGAAAGGTTCCGG 360
301 ACCGCGCGGATCAAGGACTGCAATCGGCATCGCGGTTATGCCACACGAGAAAGGTTCCGG 360
361 ATGCTCCGACTGTGATCGCCACCGAGGATGCGGTGGTGGAGGCGTCTCGCGCA 420
361 ATGCTCCGACTGTGATCGCCACCGAGGATGCGGTGGTGGAGGCGTCTCGCGCA 420
421 GGGCCACCATCGTCCGCAAGACGAACTCGAGGACATGGCATGGGATGCGGTGAGGC 480
421 GGGCCACCATCGTCCGCAAGACGAACTCGAGGACATGGCATGGGATGCGGTGAGGC 480
481 AGCGTCTACGGTCTCGCTGAAACCGCAACCGCCCGCCACCGGACCGGTTGATCTTC 540
481 AGCGTCTACGGTCTCGCTGAAACCGCAACCGCCCGCCACCGGACCGGTTGATCTTC 540
541 AGCGGCTCCGCGCTCGCTCGGCGGATGGTGGATCTGGCGCTGGCGCTCGATGAG 600
541 AGCGGCTCCGCGCTCGCTCGGCGGATGGTGGATCTGGCGCTGGCGCTCGATGAG 600
601 GCAGGACGATCCGCGATCCGCGCGCATGCTGGCGGATGAGGCGGCGGCGGCGGCGG 660
601 GCAGGACGATCCGCGATCCGCGCGCATGCTGGCGGATGAGGCGGCGGCGGCGGCGG 660
661 GGGCTGGTGGCTTACCGGCTGATACCTTCGGAGCGGACACACTGGACCACTCGGCGC 720
661 GGGCTGGTGGCTTACCGGCTGATACCTTCGGAGCGGACACACTGGACCACTCGGCGC 720
721 ATCAGGAGGCGGCTCGAGCTCAACCGCGGCTCTCGAGGTGTTGGCGGCGGCGGCGG 780
721 ATCAGGAGGCGGCTCGAGCTCAACCGCGGCTCTCGAGGTGTTGGCGGCGGCGGCGG 780
781 CGGACCTCAGTGGGTGGTAACTTCGGAGCGGAGAACTAAGGCTCGCGCTCGGCG 840
781 CGGACCTCAGTGGGTGGTAACTTCGGAGCGGAGAACTAAGGCTCGCGCTCGGCG 840
841 GAGGAGTATCCGCTCTGAGATTGCGGCTCGTGGAGGAGTCACTGGAGCGGAGCGGCG 900
841 GAGGAGTATCCGCTCTGAGATTGCGGCTCGTGGAGGAGTCACTGGAGCGGAGCGGCG 900
901 ACCGCGAGCTGATCGCGGCTTCAACGAGGAGTGGCGGCTCGAGAGCGGCGGCGG 960
901 ACCGCGAGCTGATCGCGGCTTCAACGAGGAGTGGCGGCTCGAGAGCGGCGGCGG 960
961 ACCATGAGCGGCTCGGTCGCTTGTGAGCGGCGGCTGGCTATCCAGAGCGGCGG 1020
961 ACCATGAGCGGCTCGGTCGCTTGTGAGCGGCGGCTGGCTATCCAGAGCGGCGG 1020
1021 ATGCGCTTCAACGCGCGCTATGCGGACTCCGCGGCTGTTGGGCTACTTCCACAGGG 1080
1021 ATGCGCTTCAACGCGCGCTATGCGGACTCCGCGGCTGTTGGGCTACTTCCACAGGG 1080
1081 CGGCTGAGCTGACGACCGGCTGACGACCGGCGGCGGAGTGGCGGCGGCGGAGAT 1140
1081 CGGCTGAGCTGACGACCGGCTGACGACCGGCGGCGGAGTGGCGGCGGCGGAGAT 1140
1141 CTGCGGATCTGCTCCGCGCTGATGCTGATGCTGCGGAGCACTGCGGCGGAGATACCTC 1200
1141 CTGCGGATCTGCTCCGCGCTGATGCTGATGCTGCGGAGCACTGCGGCGGAGATACCTC 1200
1201 GGCATCCACTACGGAAGGCGGAGAACTCGGCTGGAGCTCGGCAAGCATCGAGCGC 1260
1201 GGCATCCACTACGGAAGGCGGAGAACTCGGCTGGAGCTCGGCAAGCATCGAGCGC 1260
1261 GTCCTCCAGGACCGGCGTCACTGCTGACCCCGGACCAACGCGCTACCGTTGCCAAGAGCTG 1320

1261 GTCTCCAGACCGGCTGCACTGCTGACCCCGGACCGCTACCGTTGCCAAGAGCTG 1320
1321 TTGAGCGGTCGGCAGACACCATGTCCATGATCCACCGATGACGGCAATGCGATCCTC 1380
1321 TTGAGCGGTCGGCAGACACCATGTCCATGATCCACCGATGACGGCAATGCGATCCTC 1380
1381 AACACGTCGCCGCTGGACCTCACCGGTCACCCGCGCTGACGGTGGCGGCTGCGGGC 1440
1381 AACACGTCGCCGCTGGACCTCACCGGTCACCCGCGCTGACGGTGGCGGCTGCGGGC 1440
1441 GAGAGGCGCTGCCGCTGGCTCCCAAGTATAGGCGGCGGCACTTCGAGGAGTGCAGCTC 1500
1441 GAGAGGCGCTGCCGCTGGCTCCCAAGTATAGGCGGCGGCACTTCGAGGAGTGCAGCTC 1500
1501 TACCGCACCGCGCGCTGATCGAGGCGCGGCTATGGAGCTCGCGCGGAGCGGAGC 1560
1501 TACCGCACCGCGCGCTGATCGAGGCGCGGCTATGGAGCTCGCGCGGAGCGGAGC 1560
1561 GCACCGTGTGCGGTAG 1578
1561 GCACCGTGTGCGGTAG 1578

RESULT 2

US-10-156-761-2732
; Sequence 2732, Application US/10156761
; Publication No. US20030119018A1
; GENERAL INFORMATION:
; APPLICANT: OMURA, SATOSHI
; APPLICANT: IKEDA, HARUO
; APPLICANT: ISHIKAWA, JUN
; APPLICANT: HORIKAWA, HIROSHI
; APPLICANT: SHIBA, TADAYOSHI
; APPLICANT: SAKAKI, YOSHIYUKI
; APPLICANT: HATTORI, MASAHIRA
; TITLE OF INVENTION: NOVEL POLYNUCLEOTIDES
; FILE REFERENCE: 249-262
; CURRENT APPLICATION NUMBER: US/10/156,761
; CURRENT FILING DATE: 2002-05-29
; PRIOR APPLICATION NUMBER: JP 2001-204089
; PRIOR FILING DATE: 2001-05-30
; PRIOR APPLICATION NUMBER: JP 2001-272697
; PRIOR FILING DATE: 2001-08-02
; NUMBER OF SEQ ID NOS: 15109
; SEQ ID NO 2732
; LENGTH: 1503
; TYPE: DNA
; ORGANISM: Streptomyces avermitilis
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (1)..(1503)
US-10-156-761-2732

Query Match 13.2%; Score 208; DB 15; Length 1503;
Best Local Similarity 50.1%; Pred. No. 6.6e-4;
Matches 641; Conservative 0; Mismatches 615; Indels 24; Gaps 4;
QY 262 CGGTGGAGGCGCGCACGAGGCGGCGGCTGAGCGACCTGACCGCGGCGGCTCAAGGACTGC 321
DB 199 CGGAGCGCGCGGAGAGCTCGGCGGCTGGCGGCTCGCGCTCGCGCTCAAGACATC 258
QY 322 ATCGCGATTCGCGGTATGCCCAACGAAAGGTTCCGATGCTCCGACATGTCGTCGCC 381
DB 259 TTCACACGAGGCGATCCCGACCGCTCGGTTTCGAAGATCTTCGAGGGCTGATCCCG 318
QY 382 ACCGAGGATCGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGG 441
DB 319 CCTACGAGCGGACCCCTCACCAAGCTGAGGCGGCGGCGGCGGCGGCGGCGGCGG 378
QY 442 ACGAACCTCGAGGACATGGCGATGGG---TATCGGTGAAGGCGAGCGCTACGCTCTCGG 498
DB 379 ACCAACATGAGCGAGTTGCCATGGGTCTCTCCACCGAGAACAGCGGCTACGCGGCCACC 438


```
Db 3371601 --CGAGCCCGGTCGCGCGGTCGTGAGGCGCGCCGCAACGCGAGCGTCGAGGCGCATG 3371544
Qy 859 AGATTCCGGTCTCGAGGAGTCACTGAGAGCCGAAAGGTCGACGCGGACGTCGATCGCC 918
Db 3371543 CGCGTCGGCGTCTCAAGCAATTCGCGCGGAGGGCTACAGGCGGTCGTCAGCGC 3371484
Qy 919 GCGTTCAACAGGACTGCGCGGCGCTCGAGAGCGCGGTGCGACCACTGAGCGGGTCTCG 978
Db 3371483 TTGACAGTCCGTCGCGCTCTCAAGGAGCTGCGCGCGGAGATCGTCGAGTGGACTGC 3371424
Qy 979 GTGCGGTTGAGCGGCGGCTGCGCTTATCCAGAGCGCGTGTGCTTTCAACGCGCGC 1038
Db 3371423 CGGCTCTGACCTCGCGCTGTCGGGTACTACTGATCGCGCGTCCGAGTCTCTCC 3371364
Qy 1039 GCTATGCGGAGTCCGCGGTCGTGGGTACTTCCACAAGGGGCGGTGAGAGTCGACGAC 1098
Db 3371363 AACCTCGCGGCTTCGACGCGCTCGCTATGGGCTGCGGTGCGGTGAGCAGCGGAGCGC 3371304
Qy 1099 GCGTCAACGAGCGGCGGCGGAGTCGACCAACCAAGGATCTGCGGATCTGTCGCG 1158
Db 3371303 TCGCGGAGAGTCAACGCGCTCACCGGAGCGGCTTCGCGAGAGGTCAAGCGC 3371244
Qy 1159 CTGATGTCGTGATCGCGAGCACTCGCGGAGATACCTCGGCAATACCTCGCATACGCAAG 1218
Db 3371243 CGCATCTGCTCGGACGTACGCGCTCAGCTCGGCTTACTACGACGCTACTACGCGAGC 3371184
Qy 1219 GCGCAGAACCTCGCGTGGAGCTCGGCAAGCAGATCGACCGCTCTCCAGGACCGGCT 1278
Db 3371183 GCCCAGAGTCCGCTGCTCATACCGGAGCTTCGAGAGTGTTCGAGCAGGTGGAT 3371124
Qy 1279 GCACTGCTGACCGGACCGCTTACCGTTGCGCAAGAGTGTTCGAGCGTTCGCGCAAGAC 1338
Db 3371123 GTGATGCTCTCCGCGACCGCCACCGCTTCCGATCGCGGAGCGCGCGAGC 3371064
Qy 1339 ACCATGTCATGATCCACGATGACGGGCAATGCGATCTCAACGTCGCGCTGCGAC 1398
Db 3371063 CCGATGCGAGTATAC-----TCCGCGACCTGTGCACCATCCCGACCAAC 3371019
Qy 1399 CTCACCGGTACCGCGGCGTACGCGTGCACCGTGCACCGGT---GCGGCGGAGAAAGGCGCTGCC 1455
Db 3371018 CTCGCGGCAACCGCGCATGTGCTGCCCTGCGGTCTCGCGCGGAGGACACCTCCCG 3370959
Qy 1456 GTTGGCTCAGTGTAGGCGGCGCATTCGAGAGTGTGAGCGTCTTACCGCAGCGGGCC 1515
Db 3370958 GTCGCGCTGAGATCATCGCGCGCGCGCTCAAGGACGACAGGCTGTACAGGTTCGCGCC 3370899
Qy 1516 GTGATCGAGGCGCGCGCT 1535
Db 3370898 GCGGTGAGGCGCGCTTCGT 3370879
```

RESULT 4

```
US-10-282-122A-25553
; Sequence 25553, Application US/10282122A
; Publication No. US20040029129A1
; GENERAL INFORMATION:
; APPLICANT: Wang, Liangsu
; APPLICANT: Zamudio, Carlos
; APPLICANT: Malone, Cheryl
; APPLICANT: Haeelbeck, Robert
; APPLICANT: Ohlsen, Kari
; APPLICANT: Zyskind, Judith
; APPLICANT: Wall, Daniel
; APPLICANT: Trawick, John
; APPLICANT: Carr, Grant
; APPLICANT: Yamamoto, Robert
; APPLICANT: Forsyth, R.
; APPLICANT: Xu, H.
; TITLE OF INVENTION: Identification of Essential Genes in Microorganisms
; FILE REFERENCE: ELITRA.034A
; CURRENT APPLICATION NUMBER: US/10/282.122A
; CURRENT FILING DATE: 2003-02-20
```

```
; PRIOR APPLICATION NUMBER: 60/191,078
; PRIOR FILING DATE: 2000-03-21
; PRIOR APPLICATION NUMBER: 60/206,848
; PRIOR FILING DATE: 2000-05-23
; PRIOR APPLICATION NUMBER: 60/207,727
; PRIOR FILING DATE: 2000-05-26
; PRIOR APPLICATION NUMBER: 60/230,335
; PRIOR FILING DATE: 2000-09-06
; PRIOR APPLICATION NUMBER: 60/230,347
; PRIOR FILING DATE: 2000-09-09
; PRIOR APPLICATION NUMBER: 60/242,578
; PRIOR FILING DATE: 2000-10-23
; PRIOR APPLICATION NUMBER: 60/253,625
; PRIOR FILING DATE: 2000-11-27
; PRIOR APPLICATION NUMBER: 60/257,931
; PRIOR FILING DATE: 2000-12-22
; PRIOR APPLICATION NUMBER: 60/267,636
; PRIOR FILING DATE: 2001-02-09
; PRIOR APPLICATION NUMBER: 60/269,308
; PRIOR FILING DATE: 2001-02-16
; Remaining Prior Application data removed - See File Wrapper or PALM.
; NUMBER OF SEQ ID NOS: 78614
; SOFTWARE: Patent in version 3.1
; SEQ ID NO 25553
; LENGTH: 1479
; TYPE: DNA
; ORGANISM: Mycobacterium avium
; US-10-282-122A-25553

Query Match 11.1%; Score 175.2; DB 13; Length 1479;
Best Local Similarity 49.6%; Pred. No. 3.9e-32;
Matches 636; Conservative 0; Mismatches 613; Indels 33; Gaps 6;

Qy 285 GCACCTGAGCGACCTGACCGCGGATCAAGGACTGCAAGCATCGCCATCGCCATGCGCGGTATGCCAC 344
Db 213 GCGCTGCGCGGGTGCCTGCGCTGCGCTCAAGGAGCTTTCACCACTGCGATGCCAC 272
Qy 345 CACGAAACGGTCCCGATGCTCCGACTGTGATCGCCACCGAGGATCGGTGGTGGGA 404
Db 273 CACCTGCGGGTCCAAAGATTCTGCAAGGCTGCGCTGCGCTACGACGCCACCGTCAACAC 332
Qy 405 GCGGCTGTGCGGCGAGCGCGCACCATCTGCGGAAGACGAACTCGAGGACATGCGCAT 464
Db 333 GAAGCTGCGGCGCGGCGCATCCGATCTCTGGGCAAGACCAACATGAGGAGTTCGCGAT 392
Qy 465 GGGTATC--GGTGAAGGAGCGCTACGGTCTGCGTGAACCCGAAACACCCGCCCA 521
Db 393 GGGCTCTCGACCGAGAACTCGGCTACGGCGCGCAACCGCTGGAATGTGAGCG 452
Qy 522 CGGCAAGGCTGATCTTCAGCGGCTCGCGGCTGCGCTGCGCTGCGGATGTCGACTT 581
Db 453 GGTGCGCGGTTCCGCGCGGCGGCGCGGCGGCTGCGGCGGTTCCAGGCGCGCT 512
Qy 582 CGCCCTGGGCGTCTGATGAGGCGAGCATCGGATCCCGCGCGCATCGTGGGACTGCT 641
Db 513 GCGGATCGGCTCCGACACCGGCGGCTCGATCGGCGAGCGCGCGCGCTGACCGCGACCGT 572
Qy 642 CGGATGAAGGCGACCGAGCGCTGTCGCTTTACGCGCTGATACATACATGAGGACCA 701
Db 573 CGGGGTCAAGCCCACTTACGGCACCGGTGTCCGCTACGCGCTGTCGCGCTGCTATC 632
Qy 702 CTGACACACATCGGCGGCGCATCACGAGGGGGTTCGAGTCAACGCGCGGCTCTCGAGT 761
Db 633 GCTGGATCAGGGCGGCGGCTGCGCGCGCACCGTGTCTGACACCGCGCTGCTGCAAGGT 692
Qy 762 GTTGGCGGCGCGGAGCTGCGGCGGACCTCAGTGGTGGTAACTTCGCGAGCGGAGNA 821
Db 693 GATCGCGGCGGAGCATCCGCGACTCCACCTCCGT---CGACGCGCGGTCGCGCGCT 749
Qy 822 CTACGCTCCGCGCTCGGCGAGGAGTATCCGCTCTGAGATTTCGCGGTCGTGAGGAGTC 881
Db 750 CGTGGCGCGCGCGGCGGCGCGCGCGGCTGACCTCAAGGCTGTCGCGTGGGCTGCT 809
```

Qy 882 ACTGAGCCGACCGGTGCG-----ACGCCGAGCGTATCGCCCGTTCACACGAGACT 935
Db 810 CAAGCAGTGGCGCGGAGGATATCAGCCCGGCGTCTGCGCTCTTCGAGGCGCGCT 869
Qy 936 GCGCGCGCT---CGAGAGCGCGGTGCGACCATCGAGCGGTCTCGGTGCGTGTGAC 992
Db 870 CGACAGTTGACCGCGTGGCGCGGAGTCTGAGGTCTGAGTCTGCGCGCACTTCGAGTA 929
Qy 993 GCGCGCTGGCTATCCAGAGCGCGGTGAGGCTTTCAACGCGCGCGCTATGCGGACTC 1052
Db 930 CGCGTGGCGCTACTACCTGATTTTTCGCTCGGAGGTGTCAGCAACCTGGCGCGCTT 989
Qy 1053 GCGCGGTGGCTACTTCCACAGGCGCGGTGAGGTGAGCACCGCGCTCAGACGCG 1112
Db 990 CGACGCGTGGCTACTGCGGTGCGGATCGGAGTCTGAGGTGAGCACCGCGCGGAGGT 1049
Qy 1113 GCGCGAGTGGCGACACCAAGGATCTGCGGATCTGTCGCGGTGATGCTGTGAT 1172
Db 1050 GATGCGGTGACCGCGCGCGGTTCGCGCGGAGGTCAAGCGCGCATGATCGG 1109
Qy 1173 GCGGAGACCTGGCGGAGATACCTCGGATCCACTACGGAAGGCGGACCTGCG 1232
Db 1110 CACTACGCTGTCCGCGGCTACTACGACGCTTACTACACGAGCCCGCAGAGGTGCG 1169
Qy 1233 GCTGAGTGGCAAGAGATCGACGCGCTCTCCAGAGCGGCTGCACTGCTGACCCC 1292
Db 1170 CACCTGATCGCGCGGTCTCGAGCGCGCTACGAGTCTGCGAGTGTCTCGCC 1229
Qy 1293 GACACGCTTACGTTGCGCAAGAGTCTGTTAGCGGTGCGCAAGACCATGTCATGAT 1352
Db 1230 GCGCACCGCGACACCGGCTTGGGCTGGGAGGATCGACACCGCTGGCGCATGTA 1289
Qy 1353 CCCAGGATGAGGCGCAATGCGATCTCAACAGCGTCCGCTGACCTACCGGTACCC 1412
Db 1290 TCTGTGCA-----CCTGTGACGCTGCGCTGACCTGGCGCGGCGACTG 1334
Qy 1413 GCGCGTACGCTGCGGCGGAGTGGG---CGGCGAGAGGCGCTGCGCGTGGCTTCCAGT 1469
Db 1335 GCGCATGCTGCTGCGCATCGGATTTGTCGCGCGAGCAGACCTGCGGTGGGTTGCGAT 1394
Qy 1470 GATAGCGCGCACTTCGAGGAGTGCAGCTCTACCGACCGCGCGCTGATCGAGGCGCG 1529
Db 1395 CATGCGCGCGGCTGGCGAGCAGCGGTGTGACCGGTGCGCGCGCGCTACGAGGCGCG 1454
Qy 1530 CGGCTATGGAGCTCGCGCG 1551
Db 1455 GCGCGCGCGCTGCGCAGCGCG 1476

RESULT 5

US-10-282-122A-26120
; Sequence 26120, Application US/10282122A
; Publication No. US20040029129A1
; GENERAL INFORMATION:
; APPLICANT: Wang, Liangsu
; APPLICANT: Zamudio, Carlos
; APPLICANT: Malone, Cheryl
; APPLICANT: Haselbeck, Robert
; APPLICANT: Ohlsen, Kari
; APPLICANT: Zyskind, Judith
; APPLICANT: Wall, Daniel
; APPLICANT: Trawick, John
; APPLICANT: Cart, Grant
; APPLICANT: Yamamoto, Robert
; APPLICANT: Forsyth, R.
; APPLICANT: Xu, H.
; TITLE OF INVENTION: Identification of Essential Genes in Microorganisms
; FILE REFERENCE: ELITRA 034A
; CURRENT APPLICATION NUMBER: US/10/282,122A
; CURRENT FILING DATE: 2003-02-20
; PRIOR APPLICATION NUMBER: 60/191,078
; PRIOR FILING DATE: 2000-03-21
; PRIOR APPLICATION NUMBER: 60/206,848

; PRIOR FILING DATE: 2000-05-23
; PRIOR APPLICATION NUMBER: 60/207,727
; PRIOR FILING DATE: 2000-05-26
; PRIOR APPLICATION NUMBER: 60/230,335
; PRIOR FILING DATE: 2000-09-06
; PRIOR APPLICATION NUMBER: 60/230,347
; PRIOR FILING DATE: 2000-09-09
; PRIOR APPLICATION NUMBER: 60/242,578
; PRIOR FILING DATE: 2000-10-23
; PRIOR APPLICATION NUMBER: 60/253,625
; PRIOR FILING DATE: 2000-11-27
; PRIOR APPLICATION NUMBER: 60/257,931
; PRIOR FILING DATE: 2000-12-22
; PRIOR APPLICATION NUMBER: 60/267,636
; PRIOR FILING DATE: 2001-02-09
; PRIOR APPLICATION NUMBER: 60/269,308
; PRIOR FILING DATE: 2001-02-16
; Remaining Prior Application data removed - See File Wrapper or PALM.
; NUMBER OF SEQ ID NOS: 78614
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO: 26120
; LENGTH: 1482
; TYPE: DNA
; ORGANISM: Mycobacterium bovis
US-10-282-122A-26120
Query Match 10.9%; Score 172; DB 13; Length 1482;
Best Local Similarity 49.0%; Pred. No. 2,2e-31;
Matches 624; Conservative 0; Mismatches 620; Indels 30; Gaps 5;
Qy 275 CCACGAGGCGGACCTGAGCGACCTGACCGCGCGATCAAGGACTGATCGCATCGCG 334
Db 203 CTTGCGTGGCGGTGGCGGCGTCCGCTGCGCTCAAGGACGCTTCCACCCAGCG 282
Qy 335 GTATGCCCGCCAGACGCGTCCCGATGCTCCGACTGTGATCGCGACCGAGGATCGCG 394
Db 263 ACATGCCCGACCTCGGCTCAAAATCTTGAGGAGTGGGATCTCTACGACGCA 322
Qy 395 TGGTGTGAGGAGGCTGCTCGCGGAGGCGCCACCACTCTCGGCAAGACCACTCGAG 454
Db 323 CGCTGACCGCGGTTGCGCGCGCGGATCCGATCTCGGCAAGACCAACATGAGCG 382
Qy 455 ACATGCGGATGG---TATCGGTGAGGAGGAGTCTACCGTCTCGGCTGACCCGAC 511
Db 383 AGTTCGCGTGGCTCGTCAAGGAGACTTCGCTTACGCTCCACCCGCAACCGGTGA 442
Qy 512 ACCCGCCCGACGCGGTTGATCTTCCAGCGGCTCCGCGCTCGCTGCGCGGCA 571
Db 443 ATCTCGACCGGTACCGCGGTTTCCGTTGGGCGAGCGCGGCGCTTCC 502
Qy 572 TGGTGCATCTGCGCTTGGGCTCGATGAGGAGGAGTCCGGATCCCGGCGCATGCT 631
Db 503 AGGCGCGCTGGCCATCGGATCCGACACCGGCGGCTCGATCCGCGAGCGCGCGCTGA 562
Qy 632 GCGGACTGTGCGCATGAAGCGACCCACCGGCTGCTGCGCTTACGGGCTGACATACA 691
Db 563 CCGGACCGTGGGCTCAACCCACTACGCGACGCTGTCGCTATGGCTGTTGGCT 622
Qy 692 TGAACACACCTTGGACCATTCGCGGCCCATACACAGGGGGTTCGAGCTCAACCGCGG 751
Db 623 GCGGCTCTGCTGGATCAGGCGGCGCTGTCGCGGACCGCTTTTGGACACCGCGCTGT 682
Qy 752 TCCTCGAGGTGTTGGCGGCGGCGAGTGGCGGACCTCACTAGTGGTGGTAACTTCCGG 811
Db 683 TGCATCAGGTGATCGCGCGGACAGCCCGCGGCTTCCAGTCTGCTGAGGAGTGC 742
Qy 812 AGCGGAGAACTACGGCTCGCGCTCGGCGAGGAGTATCCGCTCTGAGATTCGCGTGC 871
Db 743 CCGACGTGGTGGCGCGCTAGGCGGCGCGCTCGGGGATCTCGGTGGCGTGGGCTCG 802
Qy 872 TCAGGAGTCACTGAGCGGCAACCGGTGCG-----ACGCGGAGCTGATCGCGGTTCA 925
Db 803 GCGTGGTTCACAGCTGCAAGCGGCGGAGGCTACACGCGGCGGCTGCTGGCTCTCTTCG 862

926	ACACAGGGA	CTGGCGCGCT	CGAGAGCGCC	CGGTGCGAC	CAATCGAGCGGGT	CTCGGTGCCGT	985	
863	AGGCTCCG	TGGAGAC	GTAACTAAC	CGCGTGGCGC	TGAGTCA	CGAGAGGT	CGACTGCCCG	922
986	TGTGACGG	CGCGCT	TGCGCTTAT	CCAGAGCGCG	TGATGGCTTT	CAACG---	CGCGGGCTA	1042
923	ACTTCGAC	CATGCCCT	TGGCCGCT	ATTACCTG	ATCTGATCT	CGCCCTCG	AGGTCTCGAGCAATC	982
1043	TGGCGGACT	CCGCGGTGTGG	GTACTTTC	ACAAGGGG	CGCTG	GTGAGCTC	GAGCACCGCG	1102
983	TGCGCGCT	TCGACCGG	ATCGCTAC	GGGCTG	CGGTTG	CGGTCG	CGACGCGACCGCAGCG	1042
1103	TCACGACGG	CGGCCAGAGT	TCGACCA	CCACCAC	AAGGATCT	TGGGGAT	CTCTGTCCCGGCTGA	1162
1043	CCGAGAGGT	GATGGCGAT	ACCGCGG	CCCGCTTT	CGGGCCCG	AGGTCA	AGCGGGCGCA	1102
1163	TGCTGTGT	ATCGCGGAG	CACTCG	CGACCA	GAATAC	CTCGCATC	CGCATCGCGAAGCGCGC	1222
1103	TCATGAT	TCGSCACT	TACCGT	TTCGCG	CGGCTACT	TACAGAGC	CTATTACAACAGAGCGC	1162
1223	AGAACTCG	CGGCTGGAG	CTCGGCA	AGAGAT	TCGACG	CGCTCT	CCAGGACCGGGCTGCAC	1282
1163	AGAAGT	TGCACAGCT	GATCGC	CGCACCT	TCGACG	GGGGTAT	TCGCTCGACGTGC	1222
1283	TGCTGAC	CCCGAC	CAACGCTT	ACGTTGCC	AAAGAG	GTGTTG	AGCGGTCGCGAAGACACCA	1342
1223	TGTTGT	TCGCCACA	CGACCCG	ACACCG	GTTCG	GCTTGG	GTGAGAGGTGGACGATCCGC	1282
1343	TGTCTCAT	TGATCC	CAGATG	AGCGGCA	ATCGGAT	CTCTCA	CAGTGTGCCCGCTGGA	1402
1283	TGGCGAT	TGTA	CTTGTTCGA	-----	-----	-----	CTGTGACGCTCCGCTGA	1327
1403	CCGGTACA	CCGGGCGCTG	ACGCTGCC	ACCGCTG	---CGGGCG	AGAGGGG	CTGCCCGCTTG	1459
1328	CCGGCCACT	CGGCGAT	CTGTG	TGCGT	CGGGCTCT	CCCCGG	ACGAGGTG	1387
1460	GCCTCC	AGTAGT	AGGCGG	CCCATCT	TGAGAG	GTG	CGCTTACCGCACCGGCGCGTGA	1519
1388	GCCTAC	AGATCAT	TGCGCG	CGGCA	TTCG	CCGAC	GCAGACCGGCTCTACCGGGTGGGGCGGCTT	1447
1520	TCGAGG	CGCGCG	1533					
1448	ATGAGG	CGCGCG	1461					

RESIST. 6

```

US-10-282-122A-28642
; Sequence 28642, Application US/10282122A
; Publication No. US20040029129A1
; GENERAL INFORMATION:
; APPLICANT: Wang, Liangsu
; APPLICANT: Zamudio, Carlos
; APPLICANT: Malone, Cheryl
; APPLICANT: Haselbeck, Robert
; APPLICANT: Ohlsen, Karl
; APPLICANT: Zyskind, Judith
; APPLICANT: Wall, Daniel
; APPLICANT: Trawick, John
; APPLICANT: Carr, Grant
; APPLICANT: Yamamoto, Robert
; APPLICANT: Forsyth, R.
; APPLICANT: Xu, H.
; TITLE OF INVENTION: Identification of Essential Genes in Microorganisms
; FILE REFERENCE: ELITRA.034A
; CURRENT APPLICATION NUMBER: US/10/282,122A
; CURRENT FILING DATE: 2003-02-20
; PRIOR APPLICATION NUMBER: 60/191,078
; PRIOR FILING DATE: 2000-03-21
; PRIOR APPLICATION NUMBER: 60/206,848
; PRIOR FILING DATE: 2000-05-23
; PRIOR APPLICATION NUMBER: 60/207,727
; PRIOR FILING DATE: 2000-05-26

```

	Query Match	10.8%;	Score 170.4;	DB 13;	Length 1485;
	Best Local Similarity	48.9%;	Pred. No. 5.3e-31;		
	Matches 623;	Conservative 0;	Mismatches 621;	Indels 30;	Gaps 5;
QY	275	CCACGAGGGGCACCTGAGCGACCTCACCGCCGCGATCAAGAGCTGCATCGCCATCGCCG	334		
Db	203	CTTTGCCGTGCGCGCTTGCGCGGGGTCCGCTGGCGCTCAAGAGCTGTTACCAACAGCG	262		
QY	335	GTATGCCCAACGAAACGGGTCCCGGATGCTCCCGACTGTGATCGGCACGAGATGCCG	394		
Db	263	ACATGCCCAACCACTCGCGGTCAAAAATCTTGAGGGATGGCGATCTCCCTAGCAGCCCA	322		
QY	395	TGGTGTGTGAGGGCTCTCGCGGCAGGGCCACCATCGTCGCAGAGCAAACTCGAGG	454		
Db	323	CGGTGACCGCGCGGTGTGCGCGCCGCGGGATCCCGATCTTGGGCAGAGCAAAATGAGCG	382		
QY	455	ACATGCGGATGGG---TATCGGTGAAGGAGCGAGCTCTACGGTCTCGGCTGAACCCGAAACA	511		
Db	383	AGTTTCGGATGGCTCGTTCGACGGAGAACTTCGCTTTACGTTCCACCGCAACCCGTGGA	442		
QY	512	ACCCGCGCCACGCGACGGGTGGAATCTTCAGCGGGCTCCGGCGCTGCGGTGCTGTCGCGCA	571		
Db	443	ATCTCAGCCGGTATCCCGCGGTTCCGGTGGCGGCAGCGCGGGCGGCTGCGCGCGTTCC	502		
QY	572	TGGTGCACTTCGCCCCCTGGGCGTCGATGAGCAGCGAGCATCCGCGATCCCGCGCGCATGCT	631		
Db	503	AGCGCGCGTGGCCATTCGGATTCGACACCGGGGGTTCGATCCGCGACCGCGCGCGGTGA	562		
QY	632	GCGGACTGTGCGCATGAAGGCGACCCACGGCTGTGTGCGGTCTTAAGGCGCTGACATACA	691		
Db	563	CCGCGACCGTTCGCGCTCAAAACCCACCTACGCGCACGGTGTGCGCTATGGGCTGTTGCGCT	622		
QY	692	TGGAACACACTTGGACCACTCGGCGCCATCACAGGGGGGTTCGAGCTCAACGCCCGGG	751		
Db	623	GCGCGTCTCGTGGATCATAGGCGGCGCGGTGTGCGCGCACCGTCTTGGACACCGCGCTGT	682		
QY	752	TCCTCGAGGTGTTGGCCGGGGCCGACTGGCGCGACCTCTAGTGGGTGCGTAACCTTCOCGG	811		
Db	683	TGCATCAGGTGATCGCGCGCCACGACCGCGCGACTCCACGTCCGTTCGAGCGCGAGGTGC	742		
QY	812	AGCCGAGAACTACGGCTCCGCGCTCGCGAGGGAGTATCCGCTCTGAGATTTCGCGGTG	871		
Db	743	CCACGTGTGTGGCGCGCTTAGGGCGCGCGCTTCGGGGATCTGCGTGGGCTGCGGGTGC	802		
QY	872	TCGAGGAGTCACTGGAGCCGAAACGGTGC-----ACGCCGAGCTGATCGCGCGTTCAC	925		
Db	803	GCGTGTGTCAGTAGCTTGACCGCGCGGAGGGCTACAGCCGGGCGTCTGCGCTCTCTTCG	862		
QY	926	ACCAGGAGCTGGCGCGGCTCGAGAGCGCCGTTGCGACCAATCGAGCGGGTCTCGGTGCGGT	985		

Db 863 AGGTCGCGTGGAGCAGCTAACCGCGCTGGCGCGTGGAGTCAAGGAGTCAAGTGGCGCG 922
Qy 986 TGTGACGCGCGCGCTGCGCTATCCAGAGCGCGCGTATGGTTCACG---CGCGCGCTA 1042
Db 923 ACTTCGACCATGCGCTGCGCGCTATTACCTGATTCGCGCTCGAGGTTCGAGCAATC 982
Qy 1043 TGGCGGACTCGCGCGGTGGCTACTTCCACAGGCGCGCTGGAGTCAAGCAGCGCG 1102
Db 983 TGGCGCGCTTCAGACGATGCGCTACCGGCTGCGCGTGGAGTCAAGCAGCGCG 1042
Qy 1103 TCACAGCGCGCGCGAGTCCGACACCAACAGATCTGGGATCTTCCCGGTGA 1162
Db 1043 CCGAGGAGGTGATGCGGATGACCGCGCGCGCGCTTCGCGCGCGAGTCAAGCGCGCA 1102
Qy 1163 TGCTGGTATCGCGGAGCAGCTGCGGACGATACCTCGGATCCACTACGAGAGCGC 1222
Db 1103 TCATGATCGGCACCTACGCTGTGCGCGCGCTACTAGAGCGCTATTACACGCGCG 1162
Qy 1223 AGAAGCTGCGCTGGAGTCCGCGAAGCAGATCGACGCGCTCTCCAGAGCGCGCTGCAC 1282
Db 1163 AGAAGTGGCGACGCTGATCGCGCGCGACCTCGACGCGCGGTATCGGTCCGTGACGTGC 1222
Qy 1283 TGCTGACCGCGCGCGTACCGTTCGACAGAGCTGTTCAGCGGTGCGGAGAGACCA 1342
Db 1223 TGGTGTCCCGCACGACCGCGCGCGCGCTTCGCGATGGGTGAGAGGTGAGCATCGC 1282
Qy 1343 TGTCCATGATCCACGAGTACGCGCAATGCGATCCTCAACAGCTGCGCGCTGACCTCA 1402
Db 1283 TGGCGATGATGTTGCA-----CCTGTGACGCTGCGCGTGAATTGG 1327
Qy 1403 CGGTTCACCGCGCGTACCGTGGCGCGCG---TGCGCGGAGAGCGCGCTGCGCGTTG 1459
Db 1328 CGCGCACTGCGCGATGATCTGTGCGCGTGGCGCGCTCTCCCGGAGCAGCGGTGCGCGTTG 1387
Qy 1460 GCTCCAGTATAGCGCGCACTTCGAGGAGTGCAGCTCTACCGCACGCGCGCGTGA 1519
Db 1388 GCCTACAGATGATGCGCGCGCATTTGGCGGACGCGCTCTACCGGTGGGCGCGCTT 1447
Qy 1520 TCGAGCGCGCGCGC 1533
Db 1448 ATGAGCGCGCGCGC 1461

RESULT 7

US-10-282-122A-14078
; Sequence 14078, Application US/10282122A
; Publication No. US20040029129A1
; GENERAL INFORMATION:
; APPLICANT: Wang, Liangsu
; APPLICANT: Zamudio, Carlos
; APPLICANT: Malone, Cheryl
; APPLICANT: Haselbeck, Robert
; APPLICANT: Ohlsen, Kari
; APPLICANT: Zyskind, Judith
; APPLICANT: Wall, Daniel
; APPLICANT: Trawick, John
; APPLICANT: Carr, Grant
; APPLICANT: Yamamoto, Robert
; APPLICANT: Forsyth, R.
; APPLICANT: Xu, H.
; TITLE OF INVENTION: Identification of Essential Genes in Microorganisms
; FILE REFERENCE: ELITRA.034A
; CURRENT APPLICATION NUMBER: US/10/282,122A
; CURRENT FILING DATE: 2003-02-20
; PRIOR APPLICATION NUMBER: 60/191,078
; PRIOR FILING DATE: 2000-03-21
; PRIOR APPLICATION NUMBER: 60/206,848
; PRIOR FILING DATE: 2000-05-23
; PRIOR APPLICATION NUMBER: 60/207,727
; PRIOR FILING DATE: 2000-05-26
; PRIOR APPLICATION NUMBER: 60/230,335
; PRIOR FILING DATE: 2000-09-06
; PRIOR APPLICATION NUMBER: 60/230,347

; PRIOR FILING DATE: 2000-09-09
; PRIOR APPLICATION NUMBER: 60/242,578
; PRIOR FILING DATE: 2000-10-23
; PRIOR APPLICATION NUMBER: 60/253,625
; PRIOR FILING DATE: 2000-11-27
; PRIOR APPLICATION NUMBER: 60/257,931
; PRIOR FILING DATE: 2000-12-22
; PRIOR APPLICATION NUMBER: 60/267,636
; PRIOR FILING DATE: 2001-02-09
; PRIOR APPLICATION NUMBER: 60/269,308
; PRIOR FILING DATE: 2001-02-16
; Remaining Prior Application data removed - See File Wrapper or PALM.
; NUMBER OF SEQ ID NOS: 78614
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 14078
; LENGTH: 1488
; TYPE: DNA
; ORGANISM: Burkholderia mallei
; US-10-282-122A-14078

Query Match 9.6%; Score 152.2; DB 13; Length 1488;
Best Local Similarity 48.6%; Pred. No. 1.1e-26;
Matches 631; Conservative 0; Mismatches 623; Indels 45; Gaps 6;
Qy 273 CGCCAGAGGGGCGCACTGAGCGA CCGTACCCCGCGATCAAGGACTGATCGCCATCGC 332
Db 183 CGCCAGAGGGGCGCGCTCGCGGCTCGCGATCGTGCAACAGGACGTTCGTACAGCG 242
Qy 333 CGGTATGCCACCCACAGCGGTCCCGATGTCCTCGGACTGTGATCGCCACCGAGATGC 392
Db 243 CGGCTGGCGCTCCACCGCGCGTTCGAGATGTCGCGAACTACGCGAGCCGTTGACGC 302
Qy 393 CGTGTGTGTGAGCGGCTGCTCGCGCAGCGCGCACATCGTCGGCAAGCAACCTCGA 452
Db 303 GACCGTGTGCGCGCGCTGTCGCGCGCGCGCATGTGTCAGCTCGGCAAGCAACATGA 362
Qy 453 GGCATGCGGATGGGT---ATCGGTGAAGGAGCGCTACGCTCTCGCTGAACCCGAA 509
Db 363 CGAGTTCGCGATGGGTTCGTGCAACAGAACTCCCGGCTTCGCGCGGTGAAGAACCGTG 422
Qy 510 CAACCCCGCGCCACGCGACCGGTGGATCTTTCAGCGGCTCCGCGCTGCGCTCGCTCGCG 569
Db 423 GGACACAGCGCGGTGCGCGCGCGAGCTCGCGCGCGAGCTCGCGCGCGCTCGCGCGCG 482
Qy 570 CATGTTGACTTCGCCCTGGGCTTCATGAGCGAGGAGCATCCGATCCCGCGCGCATG 629
Db 483 CCTCGCGCGCGCGACGCGCGCACACACGCGCGGCTCGATCCGCGAGCCGCGCTGTT 542
Qy 630 GTGCGGACTGGTGGCATGAAGCGCGACCCACCGCTTGGTGGCTCTTACGGCTTGACATA 689
Db 543 CGCGCGGTGACCGCGCATCAAGCCGACCTAGCGCGCGGTGTCGCGCTACGGGATGATCG 602
Qy 690 CATGAGCACACCTTGGACCATCGGCGCGCATCACAGGGGGTTCAGGTCAACCGCG 749
Db 603 GTTCGCGTGTGCTCGCTGACCGAGGCGCGCGATGGCGGCGCGCGCGGCGGATTTGCGCGCT 662
Qy 750 GGTCTCTGAGGTGTGCGCGCGCGCGACTGCGCGCGACCTTCAGTGGGT-----797
Db 663 CTGTCTGAACGCGATGGCGGCTTCAGCGAGCGGATTCGACAGGCTCGAGCGCGCGA 722
Qy 798 -----CGTTAACTTCGAGCGCGGAGAACTACGCTTCGAGGAGTCACTGAGCGCGAAACGCTCGAGCGC 845
Db 723 CGAAGACTACACGCGCGCACCTCGCGAAGGCTGCGCGCGCGCGCGCGCGCGCGCGCG 782
Qy 846 AGTATCCGCTGAGATTCGCGGTTCGTGAGGAGTCACTGAGCGCGCGAAACGCTCGAGCGC 905
Db 783 GCTCGCGCGCTCGCGCATCGGCTTCGCGCGCGGATTTTCGCGCGCGGCTTCGCGCGA---839
Qy 906 GGACGTGATCGCGCGCTTCAACAGGAGTTCGCGCGCGCTCGAGAGCGCGCGTTCGCGCAT 965
Db 840 CGACGTGCGCGCGCATCGACGCGCGCTGAAGACCTACGAGCGCTTCGCGCGCGCGCT 899
Qy 966 CGACGCGGTCTCGGTGCGCGTGTGGAGCGCGCGCTGCGCTATCCAGAGCGCGGTGATGCG 1025

Db 900 CTGCGCGTCTGCTGCGGAGAGAGCTGTGATTCCTGCTACTATGTATCGCGC 959
Qy 1026 TTTCAACGCGCGCTA---TGCGGACTCGCGCGTGTGGGTACTTCCACAAGGGGG 1082
Db 960 CGCGGAGGCTGCTCGAAGCTGTGCGCTTCGACGGCTGCTACGGCCAC-----CG 1013
Qy 1083 CGTGGAGGTGAGCAGCGCGCTCAGCAGCGGGCCAGAGTGCACACCCACAGGATCT 1142
Db 1014 TGGCGCGAATACCGGATCTGCTGACATGTAAAGAGTCCGCGCGAGGGCTTCGG 1073
Qy 1143 GGCATCTCTCCGGTGTGCTGTGATGTCGGGAGCAGCTCGCGGAGCAATACCTCGG 1202
Db 1074 GCGGAGGTGAAGCGCGGATCTCGTTCGACAGTACGTGCTGCGACGGCTACTACGA 1133
Qy 1203 CATCCACTACGAGGCGAGACCTGCGGCTGGAGCTCGGCAAGCAGATCGACCGCT 1262
Db 1134 CGCTTACTCTGAGGCGAGAGATCCCGCGGATCATCGCGAGATTTTCAGGAAGC 1193
Qy 1263 CTTCCAGGACCGGGCTGCACTGTGACCCCGACAGCGCTTACCGTTGCCAAGCAGCTGT 1322
Db 1194 GTTCAATCTCGGAGCTGATATGGGCCCGGCTCGCCACCGTTCGCTGGACATCGG 1253
Qy 1323 GAGGGTCCGAGACACCATGTCCATGATCCACGGATGACGGGCAATGCGATCCTCA 1382
Db 1254 CGGAAAGGGGAGCATCCCGTCCAGATGTATC-----TGCGGATATCTATACGCTGT 1307
Qy 1383 CAGTGCGCGCTGACCTCAGCGGTACCGCGGCTGACGTTGCCACAGGTCGCGGCGA 1442
Db 1308 GGTGAGCTCGCGGCTTCCCGGATGAGCTGCGGTGCGGCTTGGCGGGCGCGAA 1367
Qy 1443 GAGGGCTCGCGTTCAGCTCCAAAGTATGAGCGCGCACTTCGAGAGTCGACGCTCTA 1502
Db 1368 CGGAAAGCGCGGCTCGGGTGCAGATCATCGGCAACTATTTCCAGAAAGCCCGATGCT 1427
Qy 1503 CCGCACCGCGCGTGTATGAGCGCGCGCTGAGCTGCCACAGGTCGCGGCGA 1541
Db 1428 GCAGGTCCGCGAGCGGTTCAGCGCGGACCGACTGGCA 1466

RESULT 8

US-09-815-242-7924
; Sequence 7924, Application US/09815242
; Patent No. US20020061569A1
; GENERAL INFORMATION:
; APPLICANT: Haselbeck, Robert
; APPLICANT: Ohlser, Kari L.
; APPLICANT: Zyskind, Judith W.
; APPLICANT: Wall, Daniel
; APPLICANT: Trawick, John D.
; APPLICANT: Carr, Grant J.
; APPLICANT: Yamamoto, Robert T.
; APPLICANT: Xu, H. Howard
; TITLE OF INVENTION: Identification of Essential Genes in
; FILE OF INVENTION: Prokaryotes
; FILE REFERENCE: ELITRA.011A
; CURRENT APPLICATION NUMBER: US/09/815,242
; CURRENT FILING DATE: 2001-03-21
; PRIOR APPLICATION NUMBER: 60/191,078
; PRIOR FILING DATE: 2000-03-21
; PRIOR APPLICATION NUMBER: 60/206,848
; PRIOR FILING DATE: 2000-05-23
; PRIOR APPLICATION NUMBER: 60/207,727
; PRIOR FILING DATE: 2000-05-26
; PRIOR APPLICATION NUMBER: 60/242,578
; PRIOR FILING DATE: 2000-10-23
; PRIOR APPLICATION NUMBER: 60/253,625
; PRIOR FILING DATE: 2000-11-27
; PRIOR APPLICATION NUMBER: 60/257,931
; PRIOR FILING DATE: 2000-12-22
; PRIOR APPLICATION NUMBER: 60/269,308
; PRIOR FILING DATE: 2001-02-16
; NUMBER OF SEQ ID NOS: 14110

; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 7924
; LENGTH: 1455
; TYPE: DNA
; ORGANISM: Pseudomonas aeruginosa
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (1)...(1455)
US-09-815-242-7924

Query Match 9.6%; Score 151; DB 9; Length 1455;
Best Local Similarity 49.0%; Pred. No. 2.1e-26;
Matches 598; Conservative 0; Mismatches 590; Indels 33; Gaps 6;

Qy 273 CGCCACGAGGGGCACTGAGCGACCTGACCGCGCGATCAAGGACTGCTCGCATCGC 332
Db 189 CGCGAGAAACGCGCCCTGTCTCGCGCGCGATCGCCCAAGGACCTGTCTTCACCCA 248
Qy 333 CGGTATGCCACCAACGAGTCCCGATGCTCCGACTGTGATCGCCACCGAGGATGC 392
Db 249 GGGCTAGCAGCAGCTGCGGTTCGAAGATGCTCGACAATTCTGCTCGCCCTACGACG 308
Qy 393 CGTGTGTGAGCGGTGCTCGCGGAGCGCCACCATGTCGCGGAAAGCAACCTCGA 452
Db 309 CACGTGTGTGAGAGGCTCACCGCGCGCGCGTTACCTTCGCGCAAGCTGAACATGA 368
Qy 453 GGACATGCGGATGG---TATCGGTGAAGGAGGCTGTACGGTCTCTGCGTGAACCCGAA 509
Db 369 TGAATTGCCATGGCTGCTGACACCTGTCAGCCATACGCGCGGTGAAGAACCTCTG 428
Qy 510 CAACCCCGCCACGCGACGGGTGATCTTCAGCGGTTCGCGCGCTGCGGTGCTGCGG 569
Db 429 GAGCTTCACCGCGTTCGCGCGGTCTCTCGCGGTTCGCGCGGAGCTGCGCGCGCG 488
Qy 570 CATGTCGACTTCGCGCTGCGGCTGATGAGGAGGAGGAGATCGGATCCCGCGCGCATG 629
Db 489 CTGCTCGCGCGCGCACCGGACCGATACCGCGGCTGATTCGCCAACCGCGCGGCT 548
Qy 630 GTGCGACTGTTCGCGATGAAGGCGACCCAGCGCTGTGCTGCTTACGCGCTGACATA 689
Db 549 GACCAACCTCACCGGATCAAGCAACCTACCGCGCGGTTCGCGCTGGGCGATGATCGC 608
Qy 690 CATGACCAACCTTGAGCAGCATCGCGCCATCACCGAGGGGTTCGAGCTCAACGCCG 749
Db 609 CTACGCTTCCAGCTTCGACAGGCGCGCTCGCGCGCACCGCGGAGCTGCGCGCT 668
Qy 750 GGTCTCGAGGTGTGCGCGCGCGACTGCGCGAGCCCTCAGTGGGTGCGTAACCTTCC 809
Db 669 GATGCTGGGGTGTGCGCGGATTCGATCCGAAGGACTCGACACCGT-----CGAACA 722
Qy 810 GGAGCGGAGAACTACGCTCGCGCTCGCGAGGAGTATCCGCTCTGAGATTGCGCGT 869
Db 723 CGCGGTGACGACTACTTGGCGCGCTCGAAGCGCGCTGAGCGGCTTCGCGATCGGCT 782
Qy 870 CGTCGAGGAGTCACTGGAGCGGAGCGGTGCGACCGCGACGCTGATCGCGCGGTTCACCA 929
Db 783 CGCGCGGAAATATTCGCGCGCGGCTCGACAGCGCGATCGCGCGCGGTTCGCGCT 842
Qy 930 GGGACTGCGCGCTCGAGAGCGCGGTGCGACCATCGAGCGGTCTCGGTGCGGTGTG 989
Db 843 GG---TCGAGGAGCTGAAGACGCTCGCGCGCACCGTGAAGGACATTTCTTCGCCAACA 899
Qy 990 GAGCGCGGCTTGGCTATTCAGAGCGGCGTGTATGGCTTTTCAACGCGCGCTATGGCGGA 1049
Db 900 GCAGCACCCATCCCGGCTTACTAGTAATCGCGCGCGCGAGGCTCTCCACCTGTC 959
Qy 1050 TTCGCGCGGTGTGGGTACTTTCACAAGGGCGCGGTGAGCTGACGACCGCGCTCACGAC 1109
Db 960 GCGCTTCGACGCGGTGCGCTATGGTATCGTTGCGACGCGCGCGAGAACCTTGAAGACCT 1019
Qy 1110 GCGCGCGCAGAGTGCACACCCACAGGATCTCGCGATCTCTGTCGCGGTGATGCTGCT 1169
Db 1020 GTACAGCGCTCGCGCGCGGAGGCTTCGCGAGCGAGTGAAGAACCCGATCATGCTCG 1079

QY 1170 GATCGCGGAGCCTCGCGACGATACCTCGGCATCCACTACGAGCGGCGAGAACCT 1229
Db 1080 CACCTACGCACTCTCG---GCCGGCTACTAGATGCTTATTACCTGCGAGGCTCAGAAGAT 1136
QY 1230 GCGGCTGAGTCTCGCAGACGATGAGCGCTCTCCAGACCGGGCTGACCTGCTGAC 1289
Db 1137 TCGCGGCTGATCAAGAACGACTTCTGTCAGCGCTTTGCCGAAGTGAAGTCACTCTCGG 1196
QY 1290 CCGGACACGCTTACCTGTGCAACGAGCTGTGAGCGGTGCGCAAGACACCATGTCCAT 1349
Db 1197 CCGGACACGCGGACCGGCTGGAAGATCGCGGAGAGAACGACGACCGGTTTCCCA 1256
QY 1350 GATCCACGATGACGGCAATGCGATCTCAACAGTGCCCGCTGACCTACCGGTCA 1409
Db 1257 GTACCTG-----GAAGACATCTACACCATCAACCGCAACCTCGCGGCT 1301
QY 1410 CCGCGGCTGACGGTGCCACGGGTGCGGGGAGAGGGGCTGCGCGTGGCTCCAAAGT 1469
Db 1302 GCGGGGCTGTCCATGCCCGCGGCTTCTGTCG---ACGGCTGCGGCTGCGGTTCAGTT 1358
QY 1470 GATAGCGCGCACTTCGAGGA 1490
Db 1359 GCTCGCGCCTACTTCCAGGA 1379

RESULT 9

US-10-282-122A-30488
; Sequence 30488, Application US/10282122A
; Publication No. US20040029129A1
; GENERAL INFORMATION:
; APPLICANT: Wang, Lianguo
; APPLICANT: Zamudio, Carlos
; APPLICANT: Malore, Cheryl
; APPLICANT: Haselbeck, Robert
; APPLICANT: Ohlsen, Kari
; APPLICANT: Zyskind, Judith
; APPLICANT: Walli, Daniel
; APPLICANT: Trawick, John
; APPLICANT: Carr, Grant
; APPLICANT: Yamamoto, Robert
; APPLICANT: Forsyth, R.
; APPLICANT: Xu, H.
; TITLE OF INVENTION: Identification of Essential Genes in Microorganisms
; FILE REFERENCE: ELITRA.034A
; CURRENT APPLICATION NUMBER: US/10/282,122A
; CURRENT FILING DATE: 2003-02-20
; PRIOR APPLICATION NUMBER: 60/191,078
; PRIOR FILING DATE: 2000-03-21
; PRIOR APPLICATION NUMBER: 60/206,848
; PRIOR FILING DATE: 2000-05-23
; PRIOR APPLICATION NUMBER: 60/207,727
; PRIOR FILING DATE: 2000-05-26
; PRIOR APPLICATION NUMBER: 60/230,335
; PRIOR FILING DATE: 2000-09-06
; PRIOR APPLICATION NUMBER: 60/230,347
; PRIOR FILING DATE: 2000-09-09
; PRIOR APPLICATION NUMBER: 60/242,578
; PRIOR FILING DATE: 2000-10-23
; PRIOR APPLICATION NUMBER: 60/253,625
; PRIOR FILING DATE: 2000-11-27
; PRIOR APPLICATION NUMBER: 60/257,931
; PRIOR FILING DATE: 2000-12-22
; PRIOR APPLICATION NUMBER: 60/267,636
; PRIOR FILING DATE: 2001-02-09
; PRIOR APPLICATION NUMBER: 60/269,308
; PRIOR FILING DATE: 2001-02-16
; Remaining Prior Application data removed - See File Wrapper or PALM.
; NUMBER OF SEQ ID NOS: 78614
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 30488
; LENGTH: 1455
; TYPE: DNA

; ORGANISM: Pseudomonas aeruginosa

US-10-282-122A-30488

Query Match 9.6%; Score 151; DB 13; Length 1455;
Best Local Similarity 49.0%; Pred. No. 2,1e-26;
Matches 598; Conservative 0; Mismatches 590; Indels 33; Gaps 6;

QY 273 CGCCACGAGGGGCACTGAGGACCTGACCCCGCGATCAAGGACTGATCGCCATCGC 332
Db 189 CGCGAGAACGGCGCTCTCGCGCGCGGATCGCCCAAGAGACTTTCTGACCCA 248
QY 333 CGGTATGCCACACACGAGGTCCTCGGATGTCCTCGACTGTGATCGCCACGAGATGC 392
Db 249 GGGCTGACGACACAGCTGCGGTTCGAGATGTCGACAACTTCGCTCGCCCTACGACGC 308
QY 393 CGTGTGTGGAGCGGCTGCTCGCGCAGCGCGCCACCATCTGTCGGCAAGACCACTCGA 452
Db 309 CACCGTGTGCGAAGACTCACCGCGCGCGCGGTACCTCTCGGCAAGCTGAACATGGA 368
QY 453 GGACATGGCGATGGG---TATCGGTGAAGCGAGCGTCTACGGTCTCGCTGAAACCCGAA 509
Db 369 TGAATTCGCCCATGGGCTCGTGAACCAAGTCCAGCCACTACGGCGCGGTGAAGACCCCTG 428
QY 510 CAAACCCCGCCACGCGACCGGTCGATTTCCAGCGGCTCCGGCGCTCCCTCGCTCCCGG 569
Db 429 GAGCTTCGACCGGCTGCGCGGCGGCTCTCCGCGGTTCGCGCGCGCAGTCCGCGCGCG 488
QY 570 CATGTGCTGACTTGGCTTGGGCTCGATGAGGACGAGCATCCGATCCCGCGCGCATG 629
Db 489 CTTGCTGCCGCGCGCACCGGACCATACCGCGGCTCGATCCGCGCAACCGCGCGCGCT 548
QY 630 GTGCGACTGTGCTGCGCATGAAGCGGACCCACCGGCTGTGCGGCTCTTACGGCTGACATA 689
Db 549 GACCAACCTCACCGGATCAAGCAACCTTAGCGCGGTTTCCCGCTGGGCGATGATCGC 608
QY 690 CATGACACACCTTGGACACATCGGGGCCATCACAGGGGGTTCAGCTCAACCGCG 749
Db 609 CTACGCTTCCAGCTCGACACGCGCGCGCTGGCGCGCACCGCGGAGACTTCGCGCT 668
QY 750 GGTCTTCGAGGTGTGCGCGGCGGCTGCGGCGGCTCAGTGGGTGCGTAACTTCC 809
Db 669 GATGCTGGGGTGATGCGCGGATTCGATCGAAGGACTGACCGCT-----CGAACA 722
QY 810 GGAGCGGAGAACTACGGTTCGCGCTCGCGGAGGAGTATCGGTCTGAGATTCGCGGT 869
Db 723 GCGGTGGACGACTACTCTGCGCTCGAGAAAGCGCTGAGCGGCTTCGCACTCGGCT 782
QY 870 CTTGAGGAGTCACTGAGCGCGGACCGGTGCGCGCGGCGGCTGATCGCGCGCTCAACCA 929
Db 783 GCGCGGGAATACTTGGCGCGCGGCTTCGACACCGGCTCGCGGCGCGGTGCTGCGCGT 842
QY 930 GGGACTGGCGGCTCGAGAGCGCGGTCGACCATCGAGCGGCTTCGGTTCGCGTGTG 989
Db 843 GG---TCGAGGAGCTGAAGACGCTCGGCGCACGGTGAAGACATTTCCCTGCGGAACAT 899
QY 990 GAGCGGCGCTGCGCTATCCAGAGCGGCTGATGGCTTTCAACCGCGGCGCTATGCGGA 1049
Db 900 GCAGCAGCGCATCCCGGCTTACTACGTAATTCGCGCGCGGCGGAGGCTTCCTCAACCTGTC 959
QY 1050 CTCGCGCGGTGTGGGCTTACTTCACAAAGGGGCGGCTGAGCGTCAAGCACCGCGCTCAGCAC 1109
Db 960 GCGTTTCGACGCGGTGGCTATGGCTATCGTTCGAGCGCGCGCGGACCTTGAAGACCT 1019
QY 1110 GCGGCGCGGCTGCGACACCGGATCTGGGATCTCTGCGGCTGATGCTGCT 1169
Db 1020 GTACAGCGCTCGCGCGGAGGCTTCGCGAGCGAAGTGAAGAACCGGATCTATGCTCGG 1079
QY 1170 GATCGCGGACCTCGCGGACGAAATACCTCGCATCCACTACGCAAGCGCGGAGACCT 1229
Db 1080 CACTACGCACTCTCG---GCCGCTTACTAGATGCTTATCTGCGAGGCTCAGAGAT 1136
QY 1230 GCGGCTGAGCTCGGCAAGCAGATCGACGCGCTCTCCAGGACCGGCTGCACTGCTGAC 1289

Db 176 TCGCGCGCGCGCGCGCGCGCTGACCGCGCTGCCGATCGGCACACAGAGCGTGTTCG 235
QY 326 CCAATCGCCGGTATGCCACACAGAGCGGTCCCGGATGCTCCGACTGTGATCGCCACCG 385
Db 236 TACGCGCGCGCTGGCGCTCGACCGCGCTCGAAGATGCTCGGAATACGAGAGCCGCT 295
QY 386 AGGATCCGTGTGTGGAGCGCTCTCGCGCGAGCGCCACCATCGTCGGCAAGACA 445
Db 296 TCGACGCGACCGTCTGTGCGCCGCTGACGCGCGCGCATGTGTCGCTCGCAAGACA 355
QY 446 ACTCGAGACATGCGGATGGGATTCGGTGAAGGA---GCTCTACGCTCTCGCTGA 502
Db 356 ACATGACAGAGTTCGGGATGGGCTGCTCGAAGAGAAATTCGCGGTTCGGCGCGTGAAGA 415
QY 503 ACCCGAACAACCGCGCCACGCGACGCGGTGGATCTTCAGCGGCTCCGCGCTGCGCTCG 562
Db 416 ACCGTGGACACAGAGCGGTGCGCGCGCGGAGCTCGCGCGGAGCTCGCGCGCGCTCG 475
QY 563 CTGCGCGATGTCGATCTGCGCTGGCGTGCATGAGGACGACGATCGGATCCCGG 622
Db 476 CGCGCGCTCGCGCGCGCGCGACCGGACCGACACCGCGCGCTCGATCCCGACCGCG 535
QY 623 CGCATGTGCGGACTGTCGATGAGCGGACCGCGGCTGCTGCTGCTTACGCGC 682
Db 536 GTCTGTCGCGCGGTGACGCGGATCAAGCGGCTTACGCGCGCGGTGTCGCGCTACGGA 595
QY 683 TGACATATGACACACACTTGGACCATCGGCGCCATCACAGGGGGTTCGAGCTCA 742
Db 596 TGATCGCTTCGCGTCTGCTGACACGCGCGCGCGATGGCGAGCGGTCGACT 655
QY 743 AGCGCGGTCTCGAGGTGTGGCGCGGCGCGACCTGGCGCGACCGCTGATGGGT----- 797
Db 656 GCGCGTCTGCTGAACGCGATGGCGGCTTCGACGAGCGTGAACGAGCGCTCGAGC 715
QY 798 -----GCGTAACTTCCGAGCGCGAGAACTACGCTCCGCGCTCG 838
Db 716 GCGACGACGAGACTTACCGCGCCACTCGCGCGCGCTGGCGCGCGCGACGCGCG 775
QY 839 GCGAGGAGTATCGGTCTGAGATTCGCGGTGTCGAGGAGTCACTGGAGCGCAACGCTG 898
Db 776 GCAAGCGCTCGCGCGCTCGCATCGGCTTCCGCAACAGTATTTCCGCGC---CGGCC 832
QY 899 CGACGCGGACGTGATCGCGCGCTTCAACAGGAGTGGCGCGCTCGAGAGCGCGGTG 958
Db 833 TCGCGACGAGTTCGCGGACGATCGACGCGCGCTCAAGCAATACGAGCGCTCGCG 892
QY 959 CGACCATGAGCGGTCTCGGTCCGCTTGTGACGCGCGCTTCCGCTATCCAGAGCGCG 1018
Db 893 CGACGCTCGTCCGCGTCTGCTGCCAAGACGAGCTGTGATCCCGGTACTACGTGA 952
QY 1019 TGATGCTTTCACGCGCGCGCTATGGCGACTCCGCGGTGTGGCTACTTCCAGAG 1078
Db 953 TCGCGCGCGGAGCGTCTGCAACTGTGCGCTTCGACGCGGTGCGTTCGCGCCACC 1012
QY 1079 GCGCGTGGACGTGACACCGCGTCAACGCGCGCGCGCGAGTCCGACCCACCAAGG 1138
Db 1013 GCGCGCGAGTACGCGCGCTCTGACATGTACAGAGTTCGCGCGCGAGGCTTCG 1072
QY 1139 ATCTGCGATCTCTCGCGTGTATCTGATTCGCGAGCACCTGCGCGAGCAATACC 1198
Db 1073 GGCTGAAGTGAAGCGCG---GATCTGTTCGCGACGTATGTGCTGTGCGACGCTACT 1129
QY 1199 TCGGCTTCCACTACGCGAGGCGCAGAACTCGCGGTGGAGCTCGCAGAGAGATCGAG 1258
Db 1130 ACAGCGGTACTTACTGCGCGCGCAGAGATCGCGCGCATATCGGACGATTTCCAG 1189
QY 1259 CGTCTTCCAGGACCGGCTGCACTGCTGACCCCGACCGCTACCGTTCGCAACGAGC 1318
Db 1190 AAGCGTTCAGTCTTCGAGCGTATCATGGGCGCGCGCTCGCGCGATGGATC 1249
QY 1319 T 1319
Db 1250 T 1250

RESULT 12

US-10-282-122A-15096
; Sequence 15096, Application US/10282122A
; Publication No. US20040029129A1
; GENERAL INFORMATION:
; APPLICANT: Wang, Liangsu
; APPLICANT: Zamudio, Carlos
; APPLICANT: Malone, Cheryl
; APPLICANT: Haselbeck, Robert
; APPLICANT: Ohlsen, Kari
; APPLICANT: Zyskind, Judith
; APPLICANT: Wall, Daniel
; APPLICANT: Trawick, John
; APPLICANT: Carr, Grant
; APPLICANT: Yamamoto, Robert
; APPLICANT: Forsyth, R.
; APPLICANT: Xu, H.
; TITLE OF INVENTION: Identification of Essential Genes in Microorganisms
; FILE REFERENCE: ELITRA 034A
; CURRENT APPLICATION NUMBER: US/10/282,122A
; CURRENT FILING DATE: 2003-02-20
; PRIOR APPLICATION NUMBER: 60/191,078
; PRIOR FILING DATE: 2000-03-21
; PRIOR APPLICATION NUMBER: 60/206,848
; PRIOR FILING DATE: 2000-05-23
; PRIOR APPLICATION NUMBER: 60/207,727
; PRIOR FILING DATE: 2000-05-26
; PRIOR APPLICATION NUMBER: 60/230,335
; PRIOR FILING DATE: 2000-09-06
; PRIOR APPLICATION NUMBER: 60/230,347
; PRIOR FILING DATE: 2000-09-09
; PRIOR APPLICATION NUMBER: 60/242,578
; PRIOR FILING DATE: 2000-10-23
; PRIOR APPLICATION NUMBER: 60/253,625
; PRIOR FILING DATE: 2000-11-27
; PRIOR APPLICATION NUMBER: 60/257,931
; PRIOR FILING DATE: 2000-12-22
; PRIOR APPLICATION NUMBER: 60/267,636
; PRIOR FILING DATE: 2001-02-09
; PRIOR APPLICATION NUMBER: 60/269,308
; PRIOR FILING DATE: 2001-02-16
; Remaining Prior Application data removed - See File Wrapper or PALM.
; NUMBER OF SEQ ID NOS: 78614
; SOFTWARE: Patent in version 3.1
; SEQ ID NO 15096
; LENGTH: 1536
; TYPE: DNA
; ORGANISM: Bordetella pertussis
US-10-282-122A-15096

Query Match 9.0%; Score 141.4; DB 13; Length 1536;
Best Local Similarity 53.9%; Pred. No. 3.8e-24;
Matches 314; Conservative 0; Mismatches 266; Indels 3; Gaps 1;

QY 266 TGGAGGGCGCCACGAGGGGCGCACTGACGACCTGAGCGCGCGGATCAAGGATGCAATCG 325
Db 197 TGGCGCGCGGACGCGCGCGCGCTGCGGGGATCCCATCGCCCAACAGACCGCTTCG 256
QY 326 CCATCGCGGTATGCCACCAACGAGCGGTCCCGGATGCTCCCGACTGTGATCGCCACCG 385
Db 257 TCACCGTGGTGGCGCACCCACCGCAGCAGCAAGATGCTGGCGGCTATGCCAGCCGT 316
QY 386 AGGATGCGGTGCTGGAGCGGCTGTCGGGCGAGCGGCCACCATCGTGGCGAGACGA 445
Db 317 TCGACGCGCACCGTGGTTCGAGCGCTTGGCGCGCGCTGCGCGGATTCGCTGGGCAAGTGA 376
QY 446 ACCTCGAGGACATGGCGATGGGTATCGGTGAAGGACG---GTCTAGCGTCTTCGCTGA 502
Db 377 ACTCGGATGATTCGCGCATGGCTCGGCGCAACGAGACTCGGCTACGCGCGGTGCGCA 436
QY 503 ACCGAAACACCCCGCCACGCGCAGCGGATGATTTCCAGGGGCTCCGGCGCTCCCGTCG 562

D _b	437	A C C C T G G A C A C C C A G C G G T G C C C G C G C T C G T C G G C G G T C G G C C G C G G T G G C	496
Q _y	563	C T G C C G C A T G T C G A C T T C G C C C T T G G C G T C G A T G A G G C A G C A C A T C C G G A T C C C G G	622
D _b	497	C G C C C G C G T G T G T G C C G C G C A C C G S A C C G A C C G G C G G T C G G T G C G C A C C G G	556
Q _y	623	C G C A T G T G C G G A C T G G T C G G C A T G A A G G C G A C C C A C G C C T G G T G C C G T C T T A C G C C	682
D _b	557	C G C C C T G T G C G G C T C A G C G C A T C A A G C C A C C T A C G C A C G G T G T C G C G T A C G C A	616
Q _y	683	T G A C A T A C A T G A C C A C A C A C C T T G A C C A C A T C G G G C C A T C A C A G G G G G T C G A C T C A	742
D _b	617	T A A T C G C C T T C G G T C T C A C C T G N A C A G C C G C C G C T T G C G C C A G A C C G C G A C C	676
Q _y	743	A G C C C G G T C T C G A G T G T T G C C G G G G C G A C T G C G C G A C C C T C A G T G G G T G C G T A	802
D _b	677	T G C T G A A C T G C T G A C G T C A T G A C C G C T T C G A C C C C G C G A C C A C C A G C C T C A G G	736
Q _y	803	A C C T T C C G A G C C G A G A A C T A C G S C T C C G C G C T C G G C G A G G	845
D _b	737	C T G C G A C G C C A G C C C A A C A A A A C G C C C G G G T G C C C G T G G	779

RESULT 13

```

US-10-282-122A-31392
; Sequence 31392, Application US/10282122A
; Publication No. US20040029129A1
; GENERAL INFORMATION:
; APPLICANT: Wang, Liangsu
; APPLICANT: Zamudio, Carlos
; APPLICANT: Malone, Cheryl
; APPLICANT: Haselbeck, Robert
; APPLICANT: Chalsen, Karl
; APPLICANT: Zyskind, Judith
; APPLICANT: Wall, Daniel
; APPLICANT: Trawick, John
; APPLICANT: Carr, Grant
; APPLICANT: Yamamoto, Robert
; APPLICANT: Forsyth, R.
; APPLICANT: Xu, H.
; TITLE OF INVENTION: Identification of Essential Genes in Microorganisms
; FILE REFERENCE: ELITPA.034A
; CURRENT APPLICATION NUMBER: US/10/292,122A
; PRIOR FILING DATE: 2003-02-20
; PRIOR APPLICATION NUMBER: 60/191,078
; PRIOR FILING DATE: 2000-03-21
; PRIOR APPLICATION NUMBER: 60/206,849
; PRIOR FILING DATE: 2000-05-23
; PRIOR APPLICATION NUMBER: 60/207,727
; PRIOR FILING DATE: 2000-05-26
; PRIOR APPLICATION NUMBER: 60/230,335
; PRIOR FILING DATE: 2000-09-06
; PRIOR APPLICATION NUMBER: 60/230,347
; PRIOR FILING DATE: 2000-09-09
; PRIOR APPLICATION NUMBER: 60/342,578
; PRIOR FILING DATE: 2000-10-23
; PRIOR APPLICATION NUMBER: 60/253,625
; PRIOR FILING DATE: 2000-11-27
; PRIOR APPLICATION NUMBER: 60/257,931
; PRIOR FILING DATE: 2000-12-22
; PRIOR APPLICATION NUMBER: 60/267,636
; PRIOR FILING DATE: 2001-02-09
; PRIOR APPLICATION NUMBER: 60/269,308
; PRIOR FILING DATE: 2001-02-16
; Remaining Prior Application data removed - See File Wrapper or PALM.
; NUMBER OF SEQ ID NOS: 78614
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 31392
; LENGTH: 1449
; TYPE: DNA
; ORGANISM: Pseudomonas putida
US-10-282-122A-31392

```

RESULT 14
US-10-156-761-7324
; Sequence 7324, Application US/10156761
; Publication No. US20030119018A1
; GENERAL INFORMATION:
; APPLICANT: OMURA, SATOSHI
; APPLICANT: IKEDA, HARUO
; APPLICANT: ISHIKAWA, JUN
; APPLICANT: HORIKAWA, HIROSHI
; APPLICANT: SHIBA, TADAYOSHI
; APPLICANT: SAKAKI, YOSHIYUKI
; APPLICANT: HATTORI, MASAHIRA
; TITLE OF INVENTION: NOVEL POLYNUCLEOTIDES
; FILE REFERENCE: 249-262
; CURRENT APPLICATION NUMBER: US/10/156,761
; PRIOR FILING DATE: 2002-05-29
; PRIOR APPLICATION NUMBER: JP 2001-204089
; PRIOR FILING DATE: 2001-05-30
; PRIOR APPLICATION NUMBER: JP 2001-272697
; PRIOR FILING DATE: 2001-08-02
; NUMBER OF SEQ ID NOS: 15109
; SEQ ID NO 7324
; LENGTH: 1512
; TYPE: DNA
; ORGANISM: Streptomyces avermitilis
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (1)..(1512)
US-10-156-761-7324

Query Match 8.8%; Score 138.8; DB 15; Length 1512;
Best Local Similarity 52.0%; Pred. No. 1.6e-23;
Matches 386; Conservative 0; Mismatches 347; Indels 9; Gaps 3;

QY 102 CGCCAGCTGTTCCGGTCTGCGGAGATGTTGCTTGCACCTGATGACGAAT 161
DB 57 CGCCCGGTGGCGCGGCCCTGCGCAGCGAGGTGACCTCGCGGAGTCTGTCAGCG 116
QY 162 ACGCAACCGCCGCGCGCGGACCGTACACGACCGGACATCGCGCGGAAACCGAC 221
DB 117 GAGCGTGGCGGATCGAGCGGACGAGCCCTCGTGAACCGTTCGGCTGG--TGCG 173
QY 222 CGCGCAGAAAGACCGGTTCACACGATTCATCCGGTTTGTCCGGGTGGAGGGGCCACGA 281
DB 174 CGCGAGGCTGCTCGCCGAGGACACAGGCGGGGACCGGAGTGGCGCGGGCGCG 233
QY 282 GGGCCACTGAGCACTGACCGCGCGATCAAGGACTGATCGCCATCGCCGTTATGC 341
DB 234 CGGCGGCTGCTCGGGTGGCGGGTGAAGGAGCAATGGAATGCGCGGCGGAAC 293
QY 342 CACCACAAAGGGTCCCGGATGCTCCCGACTGTGATCGCCACCGAGGATCGCGTGGT 401
DB 294 GACCGCTTGGCTGTGGGGGAGTTCCCGCGCGG--CCGAGGACAGTGAAGCGGT 350
QY 402 GAGCGGCTGTCGCGGAGCGCCACATGCTGCGCAAGACGAACTCGAGACATGG- 460
DB 351 ACGCGGCTGCGGCGCGCGGGGGTGAATGCTGCGCAAGACCAACACCTGCGAGTGG 410
QY 461 --CGATGGGTATGTTGAAGCAGCGTCTACGGTCTTGGCTGAACCCGCAACACCCCG 518
DB 411 CCASTGGCGTTTACCGAGGGGCGCGTTTCGGCGCCACCGCAATCCCTGGCATCGGA 470
QY 519 CCAGGCAAGGGTGGATTTCCAGGGTCCGGCGTCCGCTGCTCGCGGCGCATGTCGA 578
DB 471 CCACAGCGCGGCGGTTCTCGGGCGGTCGCGCGCGCGGTCGCGCGGCGCTGTTGCC 530
QY 579 CTTCCGCTGGCGGTCGATGAGGAGGAGGAGGAGATCGGATCCGCGCGATGTCGAGCT 638
DB 531 GGCCCGCTTGGCTCGGACGGTTCGGTTCGTCATCCGACCTCCGACCTCTGGAAGCT 590
QY 639 GGTGCGCATGAAGCGACCCACGCGCTGTCGGTCTTACCGGCTGACATACATGGAACA 698

DB 591 GGTGCGCATCAAAACCGCAGCGGCGGATCTTCCACTGGCGCGCCCGGCTCTTCCA 650
QY 699 CACCTTGGACCAATCGGCGCCCAATACACAGGGGGGTGAGCTCAACGCGCGGGTCTCGA 758
DB 651 GGGCATCACGGTCAACGGAACGCTCGCGGTACGGTCGCGGAGCGCGCTGCTCTGGA 710
QY 759 GGTGTTGGCGGGCGGCGGACTGGCGGACCTCAGTGGTGGTAACTTTCCGAGCGCGGA 818
DB 711 CGCGGAGCGGCAACACCGGGGCGACCTGACCGCGCCCGCCCGGCTCTCTGGA 770
QY 819 GAACTACGGCTCCGCGCTCGGC 840
DB 771 GGCGTGGCGGCGAGCGCGGC 792

RESULT 15
US-10-309-294-1
; Sequence 1, Application US/10309294
; Publication No. US20030186423A1
; GENERAL INFORMATION:
; APPLICANT: VERSECK, STEFAN
; APPLICANT: DRAUZ, KARLHEINZ
; APPLICANT: BOMMARIUS, ANDREAS
; APPLICANT: KULA, MARIA-REGINA
; APPLICANT: KRIEG, LUTZ
; APPLICANT: SLUSARCZYK, HEIKE
; APPLICANT: ANSORGE-SCHUMACHER, MARION
; TITLE OF INVENTION: AMIDASE FROM VARIOVORAX
; FILE REFERENCE: 230675US
; CURRENT APPLICATION NUMBER: US/10/309,294
; PRIOR FILING DATE: 2002-12-04
; PRIOR APPLICATION NUMBER: DE 101 60 066.6
; PRIOR FILING DATE: 2001-12-06
; NUMBER OF SEQ ID NOS: 13
; SOFTWARE: Patent in version 3.1
; SEQ ID NO 1
; LENGTH: 1398
; TYPE: DNA
; ORGANISM: Variovorax paradoxus
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (1)..(1398)
; OTHER INFORMATION:
US-10-309-294-1

Query Match 7.9%; Score 125.4; DB 15; Length 1398;
Best Local Similarity 53.9%; Pred. No. 2.4e-20;
Matches 280; Conservative 0; Mismatches 236; Indels 3; Gaps 1;

QY 270 GGGCGCCACGAGGGGCGCACCTGAGCGACCTGACCGCGCGATCAAGGACTGATCGCCAT 329
DB 198 GGGCGAGATCCCGGGCGCGCTGACCGGTGCGCGGTGAAGGACCTGTGCTGGAC 257
QY 330 CGCGGCTATGCCACACGAAACGGGTCCCGATGCTCCGACTGTGATCGCCACCGAGA 389
DB 258 AAAAGGCGTGGCACGCGCGCGGATGACGCTTACCGCGACTTCGTGCCCGACGAGA 317
QY 390 TCCCGTGGTGGAGCGGCTGCTCGCGGAGCGCCACCATCGTCGCGCAAGACGAACCT 449
DB 318 CGGACGCGCGGCTGCGCAAGCTGCGGAAAGCGGTGCTGATCTCTCGGCAAGTCACT 377
QY 450 CGAGGACATGGCGATGGGTATCGGTCAAGGACGCGCTTACCGTCTGGCTGAACCCGGA 509
DB 378 CACCGAGAGCGCTATGCGGACCATCAACCCACGCT--CACGCGCGCGGTCAATCCGTG 434
QY 510 CAAACCCCGCCACGCGACCGGTGGATCTTCCAGCGGCTCCGCGGCTCGCGTCCGTCGCG 569
DB 435 GAACGCGGCGCATGGTGGGTGCTGTCGAGCGGCTCGGGCGTGGCGACCGCGCGCG 494
QY 570 CATGGTGCATTCGCGCCCTGGGGCTCGATGAGGACGAGCAGATCCGATCCCGCGCGCATG 629
DB 495 GCTTTCGTATGGCTCGCTCGGACCGACACGCGGCGGCTCGATCCGCTTCCGCTCTCGCG 554

Qy	630	GTGGGACTGGTCGGCATGAAGGGACCCACGGCCTGGTGCCGCTTACGGCCTGACATA	589
Db	555	CAACGGCCTGACCGGCTGAAGCGGACCTGGGGCCGCGTGAGCCGCGCATGGCGCCTTCGA	614
Qy	690	CATGGACACACCTTTGGACCAACATCGGGCCCATCACAGGGGGTGGAGCTCAACGCCCG	749
Db	615	GCTGGCGCCACGCTCGACCAACATCGGCCCGATGACGCGAGCGCGCGGATCGGGTGC	674
Qy	750	GGTCTCGAGGTGTTGGCCCGGGCCGACTGGCGGACCC	788
Db	675	GATGCTCGGCGCCATCGCGGAGCCGATCCGAAGGACCC	713

Search completed: June 27, 2004, 13:09:21
 Job time : 756.551 secs